Derivation for Input of Factor Graph Representation

1 Sum-Product Primal

Based on the original LP formulation

$$\begin{aligned}
\max \quad & \sum_{\alpha,\boldsymbol{x}_\alpha} b_\alpha(\boldsymbol{x}_\alpha)\theta_\alpha(\boldsymbol{x}_\alpha) + \sum_{i,x_i} b_i(x_i)\theta(x_i) \\
s.t. \quad & \forall i, \boldsymbol{b}_i \in \boldsymbol{\Delta} \\
& \forall \alpha, \boldsymbol{b}_\alpha \in \boldsymbol{\Delta} \\
& \forall i, x_i, \alpha \in N(i), \sum_{\boldsymbol{x}_\alpha \setminus x_i} b_\alpha(\boldsymbol{x}_\alpha) = b_i(x_i)
\end{aligned} \tag{1}$$

we define $\mathcal{V}^k$ as the node set allocated to the kth core. $\alpha^k_{\mathcal{F}} \in \{\alpha \cap \mathcal{V}^k \mid \forall \alpha\} \setminus \{\{i\} \mid i = 1, 2, 3..., n\}$ is defined as the index of a **free variable set** on the k-th core. Note the set of $\alpha^k_{\mathcal{F}}$ does not include single variable. Factors can share the same free variable set on the k-th core, by requiring the factors with the same free variable set to have same marginal distribution on their common free variable set, we can compress these factors and dramatically reduce dual variables to increase efficiency. In other words, we replace the original local marginal polytope with a hierarchical consistency polytope, where marginal consistency is jointly replaced by factor to free variable set and free variable set to node consistency.

$$\begin{aligned}
\max_{b} \quad & \sum_k \left\{ \begin{aligned}
& \sum_{i \in \mathcal{V}^k} \sum_{x_i} b_i(x_i)\theta_i(x_i) \\
& + \sum_{\alpha^k_{\mathcal{F}}} \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \sum_{\boldsymbol{x}_\alpha} b^k_\alpha(\boldsymbol{x}_\alpha)\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \sum_{i \in \mathcal{V}^k} \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \sum_{\boldsymbol{x}_\alpha} b^k_\alpha(\boldsymbol{x}_\alpha)\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) \\
& + \epsilon \sum_{i \in \mathcal{V}^k} c_i H_i(b_i) + \epsilon \sum_{i \in \mathcal{V}^k} \sum_{\alpha \in N^k_{\mathcal{F}}(i)} c_\alpha H^k_\alpha(b^k_\alpha) \\
& + \epsilon \sum_{\alpha^k_{\mathcal{F}}} c_{\alpha^k_{\mathcal{F}}} H^k_{\alpha^k_{\mathcal{F}}}(b^k_{\alpha^k_{\mathcal{F}}}) + \epsilon \sum_{\alpha^k_{\mathcal{F}}} \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} c_\alpha H^k_\alpha(b^k_\alpha)
\end{aligned} \right\} \\
s.t. \quad & \sum_{\boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}} b^k_\alpha(\boldsymbol{x}_\alpha) = b^k_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \qquad \forall k, \alpha^k_{\mathcal{F}}, \alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}}), \boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \\
& \sum_{\boldsymbol{x}_\alpha \setminus x_i} b^k_\alpha(\boldsymbol{x}_\alpha) = b_i(x_i) \qquad \forall k, i \in \mathcal{V}^k, x_i, \alpha \in N^k_{\mathcal{F}}(i) \\
& \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \setminus x_i} b^k_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = b_i(x_i) \qquad \forall k, \alpha^k_{\mathcal{F}}, i \in N^k(\alpha^k_{\mathcal{F}}), x_i \\
& b^k_\alpha(\boldsymbol{x}_\alpha) = b_\alpha(\boldsymbol{x}_\alpha) \qquad \forall k, \alpha : \alpha \cap \mathcal{V}^k \neq \emptyset, \boldsymbol{x}_\alpha
\end{aligned} \tag{2}$$

More formally, hierarchical consistency polytope is the feasible set of the above LP with entropy barrier functions. $\alpha^k_{\mathcal{F}}$ if the index of free variable sets on the k-th core. $\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})$ **means** α **if and only if** $\alpha \cap \mathcal{V}^k = \alpha^k_{\mathcal{F}}$**.** $i \in N^k(\alpha^k_{\mathcal{F}})$ **if and only if** i **is a neighbor of** $\alpha^k_{\mathcal{F}}$ **on the k-th core. And** $\alpha \in N^k_{\mathcal{F}}(i)$ **if and only if** $\alpha \cap \mathcal{V}^k = i$. The new formulation is a approximation to the LP relaxation of the original MAP problem. Note we distinguish factors intersecting the free variables with single variable and those with multi-variate set, so that we can have a clearer factor graph message passing representation after compressing factors.

As $\sum_{\boldsymbol{x}_\alpha \setminus x_i} b_\alpha(\boldsymbol{x}_\alpha) = \sum_{\boldsymbol{x}_\alpha \setminus x_i} b^k_\alpha(\boldsymbol{x}_\alpha) = \sum\limits_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \setminus x_i} \sum\limits_{\boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}} b^k_\alpha(\boldsymbol{x}_\alpha) = \sum\limits_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \setminus x_i} b^k_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = b_i(x_i)$, every configuration of b_α, b_i in the hierarchical consistency polytope is in the local marginal polytope.For arbitrary valid distribution denoted by b, setting $b^k_\alpha(\boldsymbol{x}_\alpha) = b(\boldsymbol{x}_\alpha)$, $b_i(x_i) = b(x_i)$ and $b^k_{\alpha_{\mathcal{F}^k}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = b(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})$, we can see the valid distribution corresponds to a point in the hierarchical consistency polytope.In other word, it is guaranteed that **the hierarchical consistency polytope (the new feasible set) is a subset of the local marginal polytope. On the other hand, every valid probability distribution in the marginal polytope is guaranteed to be in the hierarchical consistency polytope**. And we have a smaller searching space than local marginal polytope but all valid distribution in marginal polytope will be included in the new searching space.

2 Sum-Product Dual

By associating dual variables $\delta^k_\alpha(\boldsymbol{x}_{\alpha_{\mathcal{F}^k}})$, $\delta^k_\alpha(x_i)$, $\lambda^k_{i \to \alpha_{\mathcal{F}^k}}(x_i)$ and $\nu^k_\alpha(\boldsymbol{x}_\alpha)$ to the three type of constraints in Equ. 10, we have the following claim.

Claim 1. *Let $\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})$ if and only if $\alpha \cap \mathcal{V}^k = \alpha^k_{\mathcal{F}}$. The dual of the program in Equ. 10 is*

$$\min_{\lambda,\delta,\mu} \sum_k \left\{ \begin{aligned} & \sum_{i \in \mathcal{V}^k} \epsilon c_i \ln \sum_{x_i} \exp\left(\frac{\theta_i(x_i) - \sum\limits_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) - \sum\limits_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i)}{\epsilon c_i} \right) \\ & + \sum_{i \in \mathcal{V}^k} \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \epsilon c_\alpha \ln \sum_{x_i} \sum_{\boldsymbol{x}_\alpha \setminus x_i} \exp\left(\frac{\hat{\theta}_\alpha(x_i, \boldsymbol{x}_\alpha \setminus x_i) + \nu^k_\alpha(x_i, \boldsymbol{x}_\alpha \setminus x_i) + \delta^k_\alpha(x_i)}{\epsilon c_\alpha} \right) \\ & + \sum_{\alpha^k_{\mathcal{F}}} \epsilon c_{\alpha^k_{\mathcal{F}}} \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{- \sum\limits_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum\limits_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_{\alpha^k_{\mathcal{F}}}} \right) \\ & + \sum_{\alpha^k_{\mathcal{F}}} \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \sum_{\boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}, \boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}) + \nu^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}, \boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{\epsilon c_\alpha} \right) \end{aligned} \right\}$$

$$s.t \qquad \sum_{k: \alpha \cap \mathcal{V}^k \neq \emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha) = 0 \tag{3}$$

When consistency variables ν are fixed, the sub-program for the k-th core and the corresponding λ, δ update rule are shown in Claim 2

Claim 2. *When ν is fixing, the block coordinate descent can be achieve by solving the program*

$$
\begin{aligned}
&\min_{\lambda}\sum_k \left\{\begin{array}{c}\displaystyle\sum_{i\in\mathcal{V}^k}\epsilon\hat{c}_i\ln\sum_{x_i}\exp\left(\frac{\hat{\theta}_i(x_i)-\sum_{\alpha^k_{\mathcal{F}}\in N^k(i)}\lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i)}{\epsilon\hat{c}_i}\right)\\ \displaystyle+\sum_{\alpha^k_{\mathcal{F}}}\epsilon\hat{c}_{\alpha^k_{\mathcal{F}}}\ln\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}}\exp\left(\frac{\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})+\sum_{i\in N^k(\alpha^k_{\mathcal{F}})}\lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i)}{\epsilon\hat{c}_{\alpha^k_{\mathcal{F}}}}\right)\end{array}\right\}\\
&\text{where}\quad \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})=\epsilon c_\alpha\ln\sum_{\boldsymbol{x}_{\alpha\setminus\alpha^k_{\mathcal{F}}}}\exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu^k_\alpha(\boldsymbol{x}_\alpha)}{\epsilon c_\alpha}\right)\\
&\qquad\quad\ \gamma^k_\alpha(x_i)=\epsilon c_\alpha\ln\sum_{\boldsymbol{x}_\alpha\setminus x_i}\exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu^k_\alpha(\boldsymbol{x}_\alpha)}{\epsilon c_\alpha}\right)\\
&\qquad\quad\ \hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})=\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})}\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})\qquad \hat{\theta}_i(x_i)=\theta_i(x_i)+\sum_{\alpha\in N^k_{\mathcal{F}}(i)}\gamma^k_\alpha(x_i)\\
&\text{and}\qquad \hat{c}_{\alpha^k_{\mathcal{F}}}=c_{\alpha^k_{\mathcal{F}}}+\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})}c_\alpha\qquad \hat{c}_i=c_i+\sum_{\alpha\in N^k_{\mathcal{F}}(i)}c_\alpha
\end{aligned}\tag{4}
$$

3 Sum-Product Update Rule

From the program in Equ 18 and 20, the update rule is summarized in Claim 8.1.3. **Note that we use λ to represent λ^k in the following for simplicity.**

Claim 3. *The message passing rule for λ is exactly the same as convex BP rule:*

$$
\begin{aligned}
&\forall\alpha^k_{\mathcal{F}}, i\in N^k(\alpha^k_{\mathcal{F}}), x_i\quad \lambda_{i\to\alpha^k_{\mathcal{F}}}(x_i)=\frac{\hat{c}_{\alpha^k_{\mathcal{F}}}}{\tilde{c}_i}\left(\hat{\theta}_i(x_i)+\sum_{\beta^k_{\mathcal{F}}\in N^k(i)}\mu_{\beta^k_{\mathcal{F}}\to i}(x_i)\right)-\mu_{\alpha^k_{\mathcal{F}}\to i}(x_i)\\
&\text{where}\qquad \mu_{\alpha^k_{\mathcal{F}}\to i}(x_i)=\epsilon\hat{c}_{\alpha^k_{\mathcal{F}}}\ln\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}\setminus x_i}\exp\left(\frac{\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})+\sum_{j\in N^k(\alpha^k_{\mathcal{F}})\setminus i}\lambda_{j\to\alpha^k_{\mathcal{F}}}(x_j)}{\epsilon\hat{c}_{\alpha^k_{\mathcal{F}}}}\right)\\
&\text{and}\qquad \tilde{c}_i=\hat{c}_i+\sum_{\alpha^k_{\mathcal{F}}\in N^k(i)}\hat{c}_{\alpha^k_{\mathcal{F}}}
\end{aligned}\tag{5}
$$

The block coordinates rule for variables ν is

$$\nu^k_\alpha(\boldsymbol{x}_\alpha)=\frac{1}{|N_{\mathcal{P}}(\alpha)|}\sum_{j:\alpha\cap\mathcal{V}^j\neq\emptyset}\delta^j_\alpha(\boldsymbol{x}_{\alpha\cap\mathcal{V}^j})-\delta^k_\alpha(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})\tag{6}$$

where $|N_{\mathcal{P}}(\alpha)|$ is the number of sub-program in which α is involved. For arbitrary configuration of λ, variables δ can be decoded as

$$\delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})=\frac{c_\alpha}{\hat{c}_{\alpha^k_{\mathcal{F}}}}\left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})+\sum_{j\in N^k(\alpha^k_{\mathcal{F}})}\lambda_{j\to\alpha^k_{\mathcal{F}}}(x_j)\right)-\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})\tag{7}$$

and

$$\delta^k_\alpha(x_i)=\frac{c_\alpha}{\hat{c}_i}\left(\hat{\theta}_i(x_i)-\sum_{\alpha^k_{\mathcal{F}}\in N^k(i)}\lambda_{i\to\alpha^k_{\mathcal{F}}}(x_i)\right)-\gamma^k_\alpha(x_i)\tag{8}$$

which is identical to

$$\delta_\alpha^k(x_i) = \frac{c_\alpha}{\tilde{c}_i}\left(\hat{\theta}_i(x_i) + \sum_{\alpha_{\mathcal{F}}^k \in N^k(i)} \mu_{\alpha_{\mathcal{F}}^k \to i}(x_i)\right) - \gamma_\alpha^k(x_i) \tag{9}$$

by substituting λ.

4 Max-Product Primal

The LP primal formulation, which corresponds to Max-Product message passing, is

$$\begin{aligned}
\max_b \quad & \sum_k \left\{ \begin{aligned} & \sum_{i \in \mathcal{V}^k} \sum_{x_i} b_i(x_i)\theta_i(x_i) + \sum_{\alpha_{\mathcal{F}}^k} \sum_{\alpha \in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k)} \sum_{\boldsymbol{x}_\alpha} b_\alpha^k(\boldsymbol{x}_\alpha)\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) \\ & + \sum_{i \in \mathcal{V}^k} \sum_{\alpha \in N_{\mathcal{F}}^k(i)} \sum_{\boldsymbol{x}_\alpha} b_\alpha^k(\boldsymbol{x}_\alpha)\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) \end{aligned} \right\} \\
s.t. \quad & \sum_{\boldsymbol{x}_{\alpha \setminus \alpha_{\mathcal{F}}^k}} b_\alpha^k(\boldsymbol{x}_\alpha) = b_{\alpha_{\mathcal{F}}^k}^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) \qquad \forall k, \alpha_{\mathcal{F}}^k, \alpha \in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k), \boldsymbol{x}_{\alpha_{\mathcal{F}}^k} \\
& \sum_{\boldsymbol{x}_\alpha \setminus x_i} b_\alpha^k(\boldsymbol{x}_\alpha) = b_i(x_i) \qquad \forall k, i \in \mathcal{V}^k, x_i, \alpha \in N_{\mathcal{F}}^k(i) \\
& \sum_{\boldsymbol{x}_{\alpha_{\mathcal{F}}^k} \setminus x_i} b_{\alpha_{\mathcal{F}}^k}^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = b_i(x_i) \qquad \forall k, \alpha_{\mathcal{F}}^k, i \in N^k(\alpha_{\mathcal{F}}^k), x_i \\
& b_\alpha^k(\boldsymbol{x}_\alpha) = b_\alpha(\boldsymbol{x}_\alpha) \qquad \forall k, \alpha : \alpha \cap \mathcal{V}^k \neq \emptyset, \boldsymbol{x}_\alpha \\
& \sum_{x_i} b_i(x_i) = 1 \qquad \forall i \\
& \sum_{\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}} b_{\alpha_{\mathcal{F}}^k}^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = 1 \qquad \forall k, \alpha_{\mathcal{F}}^k \\
& \sum_{\boldsymbol{x}_\alpha} b_\alpha^k(\boldsymbol{x}_\alpha) = 1 \qquad \forall k, \alpha : \alpha \cap \mathcal{V}^k \neq \emptyset \\
& b_i, b_\alpha, b_\alpha^k, b_{\alpha_{\mathcal{F}}^k}^k \geq 0
\end{aligned} \tag{10}$$

5 Max-Product Dual

By associating the constraints with $\delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})$, $\delta_\alpha^k(x_i)$, $\lambda_{i \to \alpha_{\mathcal{F}}^k}^k(x_i)$, $\nu_\alpha^k(\boldsymbol{x}_\alpha)$, η_i, $\eta_{\alpha_{\mathcal{F}}^k}$, η_α^k and η_α, the dual problem is shown in the following Claim.

Claim 4. *The dual of the Max-Product problem is*

$$
\min \sum_k \left\{ \begin{array}{l} \displaystyle\sum_{i\in\mathcal{V}^k} \max_{x_i} \left(\theta_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) - \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) \right) \\ \displaystyle + \sum_{i\in\mathcal{V}^k} \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \max_{\boldsymbol{x}_\alpha} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) + \delta^k_\alpha(x_i) \right) \\ \displaystyle + \sum_{\alpha^k_{\mathcal{F}}} \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(- \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i\in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) \\ \displaystyle + \sum_{\alpha^k_{\mathcal{F}}} \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \max_{\boldsymbol{x}_\alpha} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \right) \end{array} \right\} \tag{11}
$$

$$
s.t. \quad \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha) = 0 \qquad \forall \alpha, \boldsymbol{x}_\alpha
$$

Claim 5. *When ν is fixing, the objective function of Equ 42 is lower bounded by*

$$
\min_\lambda \sum_k \left\{ \sum_{i\in\mathcal{V}^k} \max_{x_i} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) + \sum_{\alpha^k_{\mathcal{F}}} \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i\in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) \right\}
$$

$$
\begin{aligned}
\text{where} \quad & \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \max_{\boldsymbol{x}_{\alpha\setminus\alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) \right) \qquad \gamma^k_\alpha(x_i) = \max_{\boldsymbol{x}_\alpha\setminus x_i} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) \right) \\
& \hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \qquad \hat{\theta}_i(x_i) = \theta_i(x_i) + \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \gamma^k_\alpha(x_i)
\end{aligned} \tag{12}
$$

6 Max-Product Update Rule

The Max-Product Update Rule is summarized in the following. **Note that we use λ to represent λ^k in the following for simplicity.**

Claim 6. *The message passing rule for λ is exactly the same as convex BP rule:*

$$
\begin{aligned}
\forall \alpha^k_{\mathcal{F}}, i\in N^k(\alpha^k_{\mathcal{F}}), x_i \quad & \lambda_{i\to\alpha^k_{\mathcal{F}}}(x_i) = \frac{1}{1+|N^k(i)|} \left(\hat{\theta}_i(x_i) + \sum_{\beta^k_{\mathcal{F}}\in N^k(i)} \mu_{\beta^k_{\mathcal{F}}\to i}(x_i) \right) - \mu_{\alpha^k_{\mathcal{F}}\to i}(x_i) \\
\text{where} \quad & \mu_{\alpha^k_{\mathcal{F}}\to i}(x_i) = \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}\setminus x_i} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{j\in N^k(\alpha^k_{\mathcal{F}})\setminus i} \lambda_{j\to\alpha^k_{\mathcal{F}}}(x_j) \right)
\end{aligned} \tag{13}
$$

The block coordinates rule for variables ν is

$$
\nu^k_\alpha(\boldsymbol{x}_\alpha) = \frac{1}{|N_{\mathcal{P}}(\alpha)|} \sum_{j:\alpha\cap\mathcal{V}^j\neq\emptyset} \delta^j_\alpha(\boldsymbol{x}_{\alpha\cap\mathcal{V}^j}) - \delta^k_\alpha(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k}) \tag{14}
$$

where $|N_{\mathcal{P}}(\alpha)|$ is the number of sub-program in which α is involved. For arbitrary configuration of λ, variables δ can be decoded as

$$
\delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \frac{1}{1+|N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})|} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{j\in N^k(\alpha^k_{\mathcal{F}})} \lambda_{j\to\alpha^k_{\mathcal{F}}}(x_j) \right) - \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \tag{15}
$$

and

$$
\delta^k_\alpha(x_i) = \frac{1}{1+|N^k_{\mathcal{F}}(i)|} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) - \gamma^k_\alpha(x_i) \tag{16}
$$

which is identical to

$$\delta_\alpha^k(x_i) = \frac{1}{1+|N^k(i)|}\left(\hat{\theta}_i(x_i) + \sum_{\alpha_{\mathcal{F}}^k \in N^k(i)} \mu_{\alpha_{\mathcal{F}}^k \to i}(x_i)\right) - \gamma_\alpha^k(x_i) \tag{17}$$

by substituting λ.

7 Algorithm

7.1 Sum-Product

The inference algorithm can be expressed in the following sum-product fashion.

Algorithm 1 Inference

Input: $\psi_i(x_i) = \exp(\theta_i(x_i))$, $\hat{\psi}_\alpha(\boldsymbol{x}_\alpha) = \exp(\frac{\theta_\alpha(\boldsymbol{x}_\alpha)}{|N_{\mathcal{P}}(\alpha)|})$
while Until convergency **do**
  for all k **do**
    $\forall \alpha_{\mathcal{F}}^k, \alpha \in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k), \boldsymbol{x}_{\alpha_{\mathcal{F}}^k},\quad \sigma_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = \left(\sum_{\boldsymbol{x}_{\alpha \setminus \alpha_{\mathcal{F}}^k}} \left(\hat{\psi}_\alpha(\boldsymbol{x}_\alpha) n_{s\to\alpha}^k(\boldsymbol{x}_\alpha)\right)^{\frac{1}{\epsilon c_\alpha}}\right)^{\epsilon c_\alpha}$
    $\forall i \in \mathcal{V}^k, \alpha \in N_{\mathcal{F}(i)}^k, x_i,\quad \sigma_\alpha^k(x_i) = \left(\sum_{\boldsymbol{x}_\alpha \setminus x_i} \left(\hat{\psi}_\alpha(\boldsymbol{x}_\alpha) n_{s\to\alpha}^k(\boldsymbol{x}_\alpha)\right)^{\frac{1}{\epsilon c_\alpha}}\right)^{\epsilon c_\alpha}$
    $\forall \alpha_{\mathcal{F}}^k, \boldsymbol{x}_{\alpha_{\mathcal{F}}^k},\quad \hat{\psi}_{\alpha_{\mathcal{F}}^k}(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = \prod_{\alpha \in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k)} \sigma_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})$
    $\forall i \in \mathcal{V}^k, x_i,\quad \hat{\psi}_i(x_i) = \psi_i(x_i) \prod_{\alpha \in N_{\mathcal{F}}^k(i)} \sigma_\alpha^k(x_i)$
  end for
  for all k **do**
    η_α^k = Sub-Inference($\hat{\psi}_i$, $\hat{\psi}_{\alpha_{\mathcal{F}}^k}$, σ_α^k)
  end for
  for all k **do**
    $n_{s\to\alpha}^k(\boldsymbol{x}_\alpha) = \left(\prod_{j:\alpha\cap\mathcal{V}^j \neq \emptyset} n_{s\to\alpha}^j(\boldsymbol{x}_{\alpha\cap\mathcal{V}^j})\right)^{\frac{1}{|N_{\mathcal{P}}(\alpha)|}} \Big/ n_{s\to\alpha}^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})$
  end for
end while

Algorithm 2 Sub-Inference

Input: $\sigma_\alpha^k(\boldsymbol{x}_{\alpha_\mathcal{F}^k})$, $\sigma_\alpha^k(x_i)$, $\hat\psi_{\alpha_\mathcal{F}^k}(\boldsymbol{x}_{\alpha_\mathcal{F}^k})$, $\hat\psi_i(x_i)$,
$n_{i\to\alpha_\mathcal{F}^k}(x_i) = \exp(\lambda_{i\to\alpha_\mathcal{F}^k}(x_i))$, $m_{\alpha_\mathcal{F}^k\to i}(x_i) = \exp(\mu_{\alpha_\mathcal{F}^k\to i}(x_i))$
for all $t \le$ max num of inner iter **do**
  for all $i \in \mathcal{V}^k$ **do**

$$\forall \alpha_\mathcal{F}^k \in N^k(i), x_i, \quad n_{i\to\alpha_\mathcal{F}^k}(x_i) = \left(\hat\psi_i(x_i)\prod_{\beta_\mathcal{F}^k\in N^k(i)} m_{\beta_\mathcal{F}^k\to i}(x_i)\right)^{\frac{\hat c_{\alpha_\mathcal{F}^k}}{\hat c_i}} \Bigg/ m_{\alpha_\mathcal{F}^k\to i}(x_i)$$

$$\forall \alpha_\mathcal{F}^k \in N^k(i), x_i, \quad m_{\alpha_\mathcal{F}^k\to i}(x_i) = \left(\sum_{\boldsymbol{x}_{\alpha_\mathcal{F}^k}\setminus x_i}\left(\hat\psi_{\alpha_\mathcal{F}^k}(\boldsymbol{x}_{\alpha_\mathcal{F}^k})\prod_{j\in N^k(\alpha_\mathcal{F}^k)\setminus i} n_{j\to\alpha_\mathcal{F}^k}(x_j)\right)^{\frac{1}{\epsilon\hat c_{\alpha_\mathcal{F}^k}}}\right)^{\epsilon\hat c_{\alpha_\mathcal{F}^k}}$$

  end for
  for all $\alpha : \alpha \cap \mathcal{V}^k \neq \emptyset$ **do**
    if $\alpha \cap \mathcal{V}^k$ is a single node **then**

$$\eta_\alpha^k(x_i) = \left(\hat\psi_i(x_i)\prod_{\alpha_\mathcal{F}^k\in N^k(i)} m_{\alpha_\mathcal{F}^k\to i}(x_i)\right)^{\frac{c_\alpha}{\hat c_i}} \Bigg/ \sigma_\alpha^k(x_i)$$

    else

$$\eta_\alpha^k(\boldsymbol{x}_{\alpha_\mathcal{F}^k}) = \left(\hat\psi_{\alpha_\mathcal{F}^k}(\boldsymbol{x}_{\alpha_\mathcal{F}^k})\prod_{j\in N^k(\alpha_\mathcal{F}^k)} n_{j\to\alpha_\mathcal{F}^k}(x_j)\right)^{\frac{c_\alpha}{\hat c_{\alpha_\mathcal{F}^k}}} \Bigg/ \sigma_\alpha^k(\boldsymbol{x}_{\alpha_\mathcal{F}^k})$$

    end if
  end for
end for
Return η_α^k

7.2 Max-Product

The inference algorithm can be expressed in the following max-product fashion.

Algorithm 3 Inference

Input: $\theta_i(x_i)$, $\hat\theta_\alpha(\boldsymbol{x}_\alpha) = \frac{\theta_\alpha(\boldsymbol{x}_\alpha)}{|N_\mathcal{P}(\alpha)|}$
while Until convergency **do**
  for all k **do**

$$\forall \alpha_\mathcal{F}^k, \alpha \in N_\mathcal{F}^k(\alpha_\mathcal{F}^k), \boldsymbol{x}_{\alpha_\mathcal{F}^k}, \quad \gamma_\alpha^k(\boldsymbol{x}_{\alpha_\mathcal{F}^k}) = \max_{\boldsymbol{x}_{\alpha\setminus\alpha_\mathcal{F}^k}}\left(\hat\theta_\alpha(\boldsymbol{x}_\alpha) + \nu_\alpha^k(\boldsymbol{x}_\alpha)\right)$$

$$\forall i \in \mathcal{V}^k, \alpha \in N_{\mathcal{F}(i)}^k, x_i, \quad \gamma_\alpha^k(x_i) = \max_{\boldsymbol{x}_\alpha\setminus x_i}\left(\hat\theta_\alpha(\boldsymbol{x}_\alpha) + \nu_\alpha^k(\boldsymbol{x}_\alpha)\right)$$

$$\forall \alpha_\mathcal{F}^k, \boldsymbol{x}_{\alpha_\mathcal{F}^k}, \quad \hat\theta_{\alpha_\mathcal{F}^k}(\boldsymbol{x}_{\alpha_\mathcal{F}^k}) = \sum_{\alpha\in N_\mathcal{F}^k(\alpha_\mathcal{F}^k)} \gamma_\alpha^k(\boldsymbol{x}_{\alpha_\mathcal{F}^k})$$

$$\forall i \in \mathcal{V}^k, x_i, \quad \hat\theta_i(x_i) = \theta_i(x_i) + \sum_{\alpha\in N_\mathcal{F}^k(i)} \gamma_\alpha^k(x_i)$$

  end for
  for all k **do**
    δ_α^k = Sub-Inference($\hat\theta_i$, $\hat\theta_{\alpha_\mathcal{F}^k}$, γ_α^k)
  end for
  for all k **do**

$$\nu_\alpha^k(\boldsymbol{x}_\alpha) = \frac{1}{|N_\mathcal{P}(\alpha)|}\sum_{j:\alpha\cap\mathcal{V}^j\neq\emptyset} \delta_\alpha^j(\boldsymbol{x}_{\alpha\cap\mathcal{V}^j}) - \delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})$$

  end for
end while

Algorithm 4 Sub-Inference

Input: $\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}})$, $\gamma^k_\alpha(x_i)$, $\hat{\theta}_{\alpha^k_\mathcal{F}}(\boldsymbol{x}_{\alpha^k_\mathcal{F}})$, $\hat{\theta}_i(x_i)$
for all $t \le$ max num of inner iter **do**
  for all $i \in \mathcal{V}^k$ **do**
    $\forall \alpha^k_\mathcal{F} \in N^k(i), x_i, \quad \lambda_{i\to\alpha^k_\mathcal{F}}(x_i) = \frac{1}{1+|N^k(i)|}\left(\hat{\theta}_i(x_i) + \sum_{\beta^k_\mathcal{F}\in N^k(i)} \mu_{\beta^k_\mathcal{F}\to i}(x_i)\right) - \mu_{\alpha^k_\mathcal{F}\to i}(x_i)$
    $\forall \alpha^k_\mathcal{F} \in N^k(i), x_i, \quad \mu_{\alpha^k_\mathcal{F}\to i}(x_i) = \max_{\boldsymbol{x}_{\alpha^k_\mathcal{F}}\setminus x_i}\left(\hat{\theta}_{\alpha^k_\mathcal{F}}(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) + \sum_{j\in N^k(\alpha^k_\mathcal{F})\setminus i} \lambda_{j\to\alpha^k_\mathcal{F}}(x_j)\right)$
  end for
  for all $\alpha : \alpha \cap \mathcal{V}^k \neq \emptyset$ **do**
    if $\alpha \cap \mathcal{V}^k$ is a single node **then**
      $\delta^k_\alpha(x_i) = \frac{1}{1+|N^k(i)|}\left(\hat{\theta}_i(x_i) + \sum_{\alpha^k_\mathcal{F}\in N^k(i)} \mu_{\alpha^k_\mathcal{F}\to i}(x_i)\right) - \gamma^k_\alpha(x_i)$
    else
      $\delta^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) = \frac{1}{1+|N^k_\mathcal{F}(\alpha^k_\mathcal{F})|}\left(\hat{\theta}_{\alpha^k_\mathcal{F}}(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) + \sum_{j\in N^k(\alpha^k_\mathcal{F})} \lambda_{j\to\alpha^k_\mathcal{F}}(x_j)\right) - \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}})$
    end if
  end for
end for
Return δ^k_α

8 Appendix

8.1 Proof of the Claims

8.1.1 Proof of Claim 1

Claim. *Let $\alpha \in N^k_\mathcal{F}(\alpha^k_\mathcal{F})$ if and only if $\alpha \cap \mathcal{V}^k = \alpha^k_\mathcal{F}$. The dual of the program in Equ. 10 is*

$$
\min_{\lambda,\delta,\mu} \sum_k \left\{ \begin{array}{l}
\displaystyle\sum_{i\in\mathcal{V}^k} \epsilon c_i \ln \sum_{x_i} \exp\left(\frac{\theta_i(x_i) - \sum_{\alpha^k_\mathcal{F}\in N^k(i)} \lambda^k_{i\to\alpha^k_\mathcal{F}}(x_i) - \sum_{\alpha\in N^k_\mathcal{F}(i)} \delta^k_\alpha(x_i)}{\epsilon c_i}\right) \\
\displaystyle + \sum_{i\in\mathcal{V}^k} \sum_{\alpha\in N^k_\mathcal{F}(i)} \epsilon c_\alpha \ln \sum_{x_i} \sum_{\boldsymbol{x}_\alpha\setminus x_i} \exp\left(\frac{\hat{\theta}_\alpha(x_i, \boldsymbol{x}_\alpha\setminus x_i) + \nu^k_\alpha(x_i, \boldsymbol{x}_\alpha\setminus x_i) + \delta^k_\alpha(x_i)}{\epsilon c_\alpha}\right) \\
\displaystyle + \sum_{\alpha^k_\mathcal{F}} \epsilon c_{\alpha^k_\mathcal{F}} \ln \sum_{\boldsymbol{x}_{\alpha^k_\mathcal{F}}} \exp\left(\frac{-\sum_{\alpha\in N^k_\mathcal{F}(\alpha^k_\mathcal{F})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) + \sum_{i\in N^k(\alpha^k_\mathcal{F})} \lambda^k_{i\to\alpha^k_\mathcal{F}}(x_i)}{\epsilon c_{\alpha^k_\mathcal{F}}}\right) \\
\displaystyle + \sum_{\alpha^k_\mathcal{F}} \sum_{\alpha\in N^k_\mathcal{F}(\alpha^k_\mathcal{F})} \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_{\alpha^k_\mathcal{F}}} \sum_{\boldsymbol{x}_{\alpha\setminus\alpha^k_\mathcal{F}}} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}}, \boldsymbol{x}_{\alpha\setminus\alpha^k_\mathcal{F}}) + \nu^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}}, \boldsymbol{x}_{\alpha\setminus\alpha^k_\mathcal{F}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}})}{\epsilon c_\alpha}\right)
\end{array} \right\}
$$

$$
s.t \qquad \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha) = 0 \tag{18}
$$

Proof. The lagrangian dual of the primal problem is

$$
\begin{aligned}
L = \quad & \sum_k \left\{ \begin{array}{l}
\displaystyle\sum_{i\in\mathcal{V}^k}\sum_{x_i} b_i(x_i)\theta_i(x_i) + \sum_{\alpha^k_{\mathcal{F}}}\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})}\sum_{\boldsymbol{x}_\alpha} b^k_\alpha(\boldsymbol{x}_\alpha)\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) \\
\displaystyle + \sum_{i\in\mathcal{V}^k}\sum_{\alpha\in N^k_{\mathcal{F}}(i)}\sum_{\boldsymbol{x}_\alpha} b^k_\alpha(\boldsymbol{x}_\alpha)\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) \\
\displaystyle + \epsilon\sum_{i\in\mathcal{V}^k} c_i H_i(b_i) + \epsilon\sum_{\alpha^k_{\mathcal{F}}}\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} c_\alpha H^k_\alpha(b^k_\alpha) + \epsilon\sum_{i\in\mathcal{V}^k}\sum_{\alpha\in N^k_{\mathcal{F}}(i)} c_\alpha H^k_\alpha(b^k_\alpha) + \epsilon\sum_{\alpha_{\mathcal{F}^k}} c_{\alpha^k_{\mathcal{F}}} H^k_{\alpha^k_{\mathcal{F}}}(b^k_{\alpha^k_{\mathcal{F}}}) \\
\displaystyle + \sum_{\alpha^k_{\mathcal{F}}}\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})}\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \left(\sum_{\boldsymbol{x}_{\alpha\setminus\alpha^k_{\mathcal{F}}}} b^k_\alpha(\boldsymbol{x}_\alpha) - b^k_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \right) \\
\displaystyle + \sum_{i\in\mathcal{V}^k}\sum_{\alpha\in N^k_{\mathcal{F}}(i)}\sum_{x_i} \delta^k_\alpha(x_i) \left(\sum_{\boldsymbol{x}_\alpha\setminus x_i} b^k_\alpha(\boldsymbol{x}_\alpha) - b_i(x_i) \right) \\
\displaystyle + \sum_{\alpha^k_{\mathcal{F}}}\sum_{i\in\mathcal{V}^k}\sum_{x_i} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \left(\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}\setminus x_i} b^k_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) - b_i(x_i) \right) \\
\displaystyle + \sum_{\alpha^k_{\mathcal{F}}}\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})}\sum_{\boldsymbol{x}_\alpha} \nu^k_\alpha(\boldsymbol{x}_\alpha)\Big(b^k_\alpha(\boldsymbol{x}_\alpha) - b_\alpha(\boldsymbol{x}_\alpha)\Big) \\
\displaystyle + \sum_{i\in\mathcal{V}^k}\sum_{\alpha\in N^k_{\mathcal{F}}(i)}\sum_{\boldsymbol{x}_\alpha} \nu^k_\alpha(\boldsymbol{x}_\alpha)\Big(b^k_\alpha(\boldsymbol{x}_\alpha) - b_\alpha(\boldsymbol{x}_\alpha)\Big)
\end{array} \right\} \\
= \quad & \sum_k \left\{ \begin{array}{l}
\displaystyle\sum_{i\in\mathcal{V}^k}\sum_{x_i} b_i(x_i) \left(\theta_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) - \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) \right) + \epsilon\sum_{i\in\mathcal{V}^k} c_i H_i(b_i) \\
\displaystyle + \sum_{i\in\mathcal{V}^k}\sum_{\alpha\in N^k_{\mathcal{F}}(i)}\sum_{\boldsymbol{x}_\alpha} b^k_\alpha(\boldsymbol{x}_\alpha) \left(\hat{\theta}_\alpha(x_i, \boldsymbol{x}_\alpha\setminus x_i) + \nu^k_\alpha(x_i, \boldsymbol{x}_\alpha\setminus x_i) + \delta^k_\alpha(x_i) \right) \\
\displaystyle + \epsilon\sum_{i\in\mathcal{V}^k}\sum_{\alpha\in N^k_{\mathcal{F}}(i)} c_\alpha H^k_\alpha(b^k_\alpha) \\
\displaystyle + \sum_{\alpha^k_{\mathcal{F}}}\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} b^k_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha_{\mathcal{F}^k}}) \left(- \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i\in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) \\
\displaystyle + \epsilon\sum_{\alpha^k_{\mathcal{F}}} c_{\alpha^k_{\mathcal{F}}} H^k_{\alpha^k_{\mathcal{F}}}(b^k_{\alpha^k_{\mathcal{F}}}) \\
\displaystyle + \sum_{\alpha^k_{\mathcal{F}}}\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})}\sum_{\boldsymbol{x}_\alpha} b^k_\alpha(\boldsymbol{x}_\alpha) \left(\hat{\theta}_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}, \boldsymbol{x}_{\alpha\setminus\alpha^k_{\mathcal{F}}}) + \nu^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}, \boldsymbol{x}_{\alpha\setminus\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \right) \\
\displaystyle + \epsilon\sum_{\alpha^k_{\mathcal{F}}}\sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} c_\alpha H^k_\alpha(b^k_\alpha)
\end{array} \right\} \\
- \quad & \sum_\alpha\sum_{\boldsymbol{x}_\alpha} b_\alpha(\boldsymbol{x}_\alpha) \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha)
\end{aligned}
\tag{19}
$$

We maximize L over b analytically with the fact that log-sum-exp function is the conjugate function of entropy under the simplex constraints. In addition, to prevent $\sup_{b_\alpha(\boldsymbol{x}_\alpha)} b_\alpha(\boldsymbol{x}_\alpha) \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha)$ from going to negative infinity, the last term in the lagrangian dual gives additional constraints of $\sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha) = 0$, which results in the dual program in Claim 1 □

8.1.2 Proof of Claim 2

Claim. *When ν is fixing, the block coordinate descent can be achieve by solving the program*

$$\min_{\lambda} \sum_k \left\{ \begin{array}{l} \displaystyle\sum_{i \in \mathcal{V}^k} \epsilon \hat{c}_i \ln \sum_{x_i} \exp\left(\frac{\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon \hat{c}_i} \right) \\ \displaystyle + \sum_{\alpha^k_{\mathcal{F}}} \epsilon \hat{c}_{\alpha^k_{\mathcal{F}}} \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon \hat{c}_{\alpha^k_{\mathcal{F}}}} \right) \end{array} \right\}$$

$$\begin{aligned} \text{where} \quad & \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha)}{\epsilon c_\alpha} \right) \\ & \gamma^k_\alpha(x_i) = \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_\alpha \setminus x_i} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha)}{\epsilon c_\alpha} \right) \\ & \hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \qquad \hat{\theta}_i(x_i) = \theta_i(x_i) + \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \gamma^k_\alpha(x_i) \\ \text{and} \quad & \hat{c}_{\alpha^k_{\mathcal{F}}} = c_{\alpha^k_{\mathcal{F}}} + \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} c_\alpha \qquad \hat{c}_i = c_i + \sum_{\alpha \in N^k_{\mathcal{F}}(i)} c_\alpha \end{aligned} \tag{20}$$

Proof. **Derivation for free var set terms:**

By extracting all the terms involving $\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}$ from Equ 18, we have

$$\begin{aligned} L_{\alpha^k_{\mathcal{F}}} =& \epsilon c_{\alpha^k_{\mathcal{F}}} \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_{\alpha^k_{\mathcal{F}}}} \right) \\ & + \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \sum_{\boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}, \boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}) + \nu^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}, \boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{\epsilon c_\alpha} \right) \\ =& \epsilon c_{\alpha^k_{\mathcal{F}}} \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_{\alpha^k_{\mathcal{F}}}} \right) \\ & + \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{\epsilon c_\alpha} \right) \end{aligned} \tag{21}$$

Setting the derivative with respect to $\delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})$ to 0, we have

$$\frac{\exp\left(\frac{\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{\epsilon c_\alpha} \right)}{\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{\epsilon c_\alpha} \right)} = \frac{\exp\left(\frac{- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_{\alpha^k_{\mathcal{F}}}} \right)}{\sum_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_{\alpha^k_{\mathcal{F}}}} \right)} \tag{22}$$

Introducing a degree of freedom in normalization, we have

$$\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \frac{c_\alpha}{c_{\alpha^k_{\mathcal{F}}}} \left(- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) \tag{23}$$

Summing over $\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})$, it gives

$$\sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \frac{\sum\limits_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} c_\alpha}{\hat{c}_{\alpha^k_{\mathcal{F}}}} \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) - \frac{c_{\alpha^k_{\mathcal{F}}}}{\hat{c}_{\alpha^k_{\mathcal{F}}}} \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \tag{24}$$

Substituting it back to Equ 9, we have

$$\frac{\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{c_\alpha} = \frac{-\sum\limits_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum\limits_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{c_{\alpha^k_{\mathcal{F}}}} = \frac{\sum\limits_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) + \sum\limits_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}})}{\hat{c}_{\alpha^k_{\mathcal{F}}}} \tag{25}$$

with which we compress $L_{\alpha^k_{\mathcal{F}}}$ in Equ 21 into a single term

$$L_{\alpha^k_{\mathcal{F}}} = \epsilon \hat{c}_{\alpha^k_{\mathcal{F}}} \ln \left(\frac{\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum\limits_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon \hat{c}_{\alpha^k_{\mathcal{F}}}} \right) \tag{26}$$

Derivation for node terms:

The derivation goes similarly to the above process for terms which are compressed into node. We also put it here for clearer future reference.

$$\begin{aligned}
L_i =& \epsilon c_i \ln \sum_{x_i} \exp \left(\frac{\theta_i(x_i) - \sum\limits_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) - \sum\limits_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i)}{\epsilon c_i} \right) \\
&+ \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \epsilon c_\alpha \ln \sum_{x_i} \sum_{\boldsymbol{x}_\alpha \backslash x_i} \exp \left(\frac{\hat{\theta}_\alpha(x_i, \boldsymbol{x}_\alpha \backslash x_i) + \nu^k_\alpha(x_i, \boldsymbol{x}_\alpha \backslash x_i) + \delta^k_\alpha(x_i)}{\epsilon c_\alpha} \right) \\
=& \epsilon c_i \ln \sum_{x_i} \exp \left(\frac{\theta_i(x_i) - \sum\limits_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) - \sum\limits_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i)}{\epsilon c_{\alpha^k_{\mathcal{F}}}} \right) \\
&+ \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \epsilon c_\alpha \ln \sum_{x_i} \exp \left(\frac{\gamma^k_\alpha(x_i) + \delta^k_\alpha(x_i)}{\epsilon c_\alpha} \right)
\end{aligned} \tag{27}$$

Setting the derivative with respect to $\delta^k_\alpha(x_i)$ to 0, we have

$$\frac{\exp\left(\frac{\gamma^k_\alpha(x_i) + \delta^k_\alpha(x_i)}{\epsilon c_\alpha}\right)}{\sum\limits_{x_i} \exp\left(\frac{\gamma^k_\alpha(x_i) + \delta^k_\alpha(x_i)}{\epsilon c_\alpha}\right)} = \frac{\exp\left(\frac{\theta_i(x_i) - \sum\limits_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) - \sum\limits_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_i}\right)}{\sum\limits_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \exp\left(\frac{\theta_i(x_i) - \sum\limits_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) - \sum\limits_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i)}{\epsilon c_i}\right)} \tag{28}$$

Introducing a degree of freedom in normalization, we have

$$\gamma^k_\alpha(x_i) + \delta^k_\alpha(x_i) = \frac{c_\alpha}{c_i} \left(\theta_i(x_i) - \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) - \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) \tag{29}$$

Summing over $\alpha \in N^k_{\mathcal{F}}(i)$, it gives

$$\sum_{\alpha \in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) = \frac{\sum\limits_{\alpha \in N^k_{\mathcal{F}}(i)} c_\alpha}{\hat{c}_i} \left(\theta_i(x_i) - \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) - \frac{c_i}{\hat{c}_i} \sum_{\alpha \in N^k_{\mathcal{F}}(i)} \gamma^k_\alpha(x_i) \tag{30}$$

Substituting it back to Equ 29, we have

$$\frac{\gamma_\alpha^k(x_i)+\delta_\alpha^k(x_i)}{c_\alpha} = \frac{\theta_i(x_i) - \sum_{\alpha\in N_{\mathcal{F}}^k(i)} \delta_\alpha^k(x_i) - \sum_{\alpha_{\mathcal{F}}^k\in N^k(i)} \lambda_{i\to\alpha_{\mathcal{F}}^k}^k(x_i)}{c_i} = \frac{\left(\theta_i(x_i) + \sum_{\alpha\in N_{\mathcal{F}}^k(i)} \gamma_\alpha^k(x_i)\right) - \sum_{\alpha_{\mathcal{F}}^k\in N^k(i)} \lambda_{i\to\alpha_{\mathcal{F}}^k}^k(x_i)}{\hat{c}_i} \tag{31}$$

with which we compress L_i in Equ 21 into a single term

$$L_i = \epsilon\hat{c}_i \ln\sum_{x_i}\exp\left(\frac{\hat{\theta}_i(x_i) - \sum_{\alpha_{\mathcal{F}}^k\in N^k(i)} \lambda_{i\to\alpha_{\mathcal{F}}^k}^k(x_i)}{\epsilon\hat{c}_i}\right) \tag{32}$$

Combining Equ 26 and Equ 32 will give us the expression in Equ 20. □

8.1.3 Proof of Claim 3

Claim. *The message passing rule for λ is exactly the same as convex BP rule:*

$$\begin{aligned}
&\forall\alpha_{\mathcal{F}}^k, i\in N^k(\alpha_{\mathcal{F}}^k), x_i \quad && \lambda_{i\to\alpha_{\mathcal{F}}^k}(x_i) = \frac{\hat{c}_{\alpha_{\mathcal{F}}^k}}{\tilde{c}_i}\left(\hat{\theta}_i(x_i) + \sum_{\beta_{\mathcal{F}}^k\in N^k(i)} \mu_{\beta_{\mathcal{F}}^k\to i}(x_i)\right) - \mu_{\alpha_{\mathcal{F}}^k\to i}(x_i)\\
&\text{where} && \mu_{\alpha_{\mathcal{F}}^k\to i}(x_i) = \epsilon\hat{c}_{\alpha_{\mathcal{F}}^k}\ln\sum_{\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}\setminus x_i}\exp\left(\frac{\hat{\theta}_{\alpha_{\mathcal{F}}^k}(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) + \sum_{j\in N^k(\alpha_{\mathcal{F}}^k)\setminus i}\lambda_{j\to\alpha_{\mathcal{F}}^k}(x_j)}{\epsilon\hat{c}_{\alpha_{\mathcal{F}}^k}}\right)\\
&\text{and} && \tilde{c}_i = \hat{c}_i + \sum_{\alpha_{\mathcal{F}}^k\in N^k(i)}\hat{c}_{\alpha_{\mathcal{F}}^k}
\end{aligned} \tag{33}$$

The block coordinates rule for variables ν is

$$\nu_\alpha^k(\boldsymbol{x}_\alpha) = \frac{1}{|N_{\mathcal{P}}(\alpha)|}\sum_{j:\alpha\cap\mathcal{V}^j\neq\emptyset}\delta_\alpha^j(\boldsymbol{x}_{\alpha\cap\mathcal{V}^j}) - \delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k}) \tag{34}$$

where $|N_{\mathcal{P}}(\alpha)|$ is the number of sub-program in which α is involved. For arbitrary configuration of λ, variables δ can be decoded as

$$\delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = \frac{c_\alpha}{\hat{c}_{\alpha_{\mathcal{F}}^k}}\left(\hat{\theta}_{\alpha_{\mathcal{F}}^k}(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) + \sum_{j\in N^k(\alpha_{\mathcal{F}}^k)}\lambda_{j\to\alpha_{\mathcal{F}}^k}(x_j)\right) - \gamma_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) \tag{35}$$

and

$$\delta_\alpha^k(x_i) = \frac{c_\alpha}{\hat{c}_i}\left(\hat{\theta}_i(x_i) - \sum_{\alpha_{\mathcal{F}}^k\in N^k(i)}\lambda_{i\to\alpha_{\mathcal{F}}^k}(x_i)\right) - \gamma_\alpha^k(x_i) \tag{36}$$

which is identical to

$$\delta_\alpha^k(x_i) = \frac{c_\alpha}{\tilde{c}_i}\left(\hat{\theta}_i(x_i) + \sum_{\alpha_{\mathcal{F}}^k\in N^k(i)}\mu_{\alpha_{\mathcal{F}}^k\to i}(x_i)\right) - \gamma_\alpha^k(x_i) \tag{37}$$

by substituting λ.

Proof. For λ variables, the rule is exactly the same as convex BP.

In terms of ν, from the program in Claim 1, we take the lagrangian dual of the objective function

$$\hat{L}_\alpha = \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_\alpha} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu_\alpha^k(\boldsymbol{x}_\alpha) + \delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})}{\epsilon c_\alpha}\right) + \sum_{\boldsymbol{x}_\alpha} \eta_\alpha(\boldsymbol{x}_\alpha) \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu_\alpha^k(\boldsymbol{x}_\alpha) \tag{38}$$

setting the derivative with respect to $\nu_\alpha^k(\boldsymbol{x}_\alpha)$ as zero, which gives the following equation

$$\frac{\exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu_\alpha^k(\boldsymbol{x}_\alpha)+\delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})}{\epsilon c_\alpha}\right)}{\sum_{\boldsymbol{x}_\alpha}\exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu_\alpha^k(\boldsymbol{x}_\alpha)+\delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})}{\epsilon c_\alpha}\right)} = \eta_\alpha(\boldsymbol{x}_\alpha) \tag{39}$$

As any set of $\nu_\alpha^k(\boldsymbol{x}_\alpha)$ satisfying the above equation is guaranteed to be optimal and every optimal is guaranteed to satisfy $\sum\limits_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu_\alpha^k(\boldsymbol{x}_\alpha) = 0$, we can add arbitrary additional constant to $\nu_\alpha^k(\boldsymbol{x}_\alpha)$ as long as Equ 39 is satisfied. The effect of additional constant is accumulated to 0 over k . Thus the resulting optimal value for $\hat{L}_\alpha$ is unchanged. i.e. we can simply take

$$\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu_\alpha^k(\boldsymbol{x}_\alpha) + \delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k}) = \beta_\alpha(\boldsymbol{x}_\alpha) \tag{40}$$

summing over $k : \alpha \cap \mathcal{V}^k \neq \emptyset$, $\nu_\alpha^k(\boldsymbol{x}_\alpha)$ are eliminated which gives

$$\theta_\alpha(\boldsymbol{x}_\alpha) + \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k}) = |N_{\mathcal{P}}(\alpha)|\beta_\alpha(\boldsymbol{x}_\alpha) \tag{41}$$

We can substituting it back to Equ 40 and get the rule in Claim 8.1.3. Note the explicit expression of δ in Equ 35 and Equ 36 can be directly derived from Equ 25 and 31. □

8.1.4 Proof of Claim 4

Claim. *The dual of the Max-Product problem is*

$$\begin{aligned}
\min \quad & \sum_k \left\{ \begin{array}{l}
\sum\limits_{i\in\mathcal{V}^k} \max\limits_{x_i} \left(\theta_i(x_i) - \sum\limits_{\alpha_{\mathcal{F}}^k\in N^k(i)} \lambda^k_{i\to\alpha_{\mathcal{F}}^k}(x_i) - \sum\limits_{\alpha\in N_{\mathcal{F}}^k(i)} \delta_\alpha^k(x_i)\right) \\
+ \sum\limits_{i\in\mathcal{V}^k} \sum\limits_{\alpha\in N_{\mathcal{F}}^k(i)} \max\limits_{\boldsymbol{x}_\alpha} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu_\alpha^k(\boldsymbol{x}_\alpha) + \delta_\alpha^k(x_i)\right) \\
+ \sum\limits_{\alpha_{\mathcal{F}}^k} \max\limits_{\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}} \left(- \sum\limits_{\alpha\in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k)} \delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) + \sum\limits_{i\in N^k(\alpha_{\mathcal{F}}^k)} \lambda^k_{i\to\alpha_{\mathcal{F}}^k}(x_i)\right) \\
+ \sum\limits_{\alpha_{\mathcal{F}}^k} \sum\limits_{\alpha\in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k)} \max\limits_{\boldsymbol{x}_\alpha} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu_\alpha^k(\boldsymbol{x}_\alpha) + \delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})\right)
\end{array} \right\} \\
s.t. \quad & \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu_\alpha^k(\boldsymbol{x}_\alpha) = 0 \qquad \forall \alpha, \boldsymbol{x}_\alpha
\end{aligned} \tag{42}$$

Proof. With standard LP primal-dual tranformation

$$\begin{array}{llcll}
\max & \boldsymbol{c}^T\boldsymbol{x} & = & \min & \boldsymbol{b}^T\boldsymbol{y} \\
s.t. & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} & & s.t. & \boldsymbol{A}^T\boldsymbol{y} \geq \boldsymbol{c} \\
 & \boldsymbol{x} \geq 0 & & &
\end{array} \tag{43}$$

we have the following dual formulation

$$
\begin{aligned}
\min \quad & \sum_k \left\{ \sum_{i\in\mathcal{V}^k} \eta_i(x_i) + \sum_{i\in\mathcal{V}^k} \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \eta^k_\alpha + \sum_{\alpha^k_{\mathcal{F}}} \eta^k_{\alpha^k_{\mathcal{F}}} + \sum_{\alpha^k_{\mathcal{F}}} \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \eta^k_\alpha \right\} \\
s.t. \quad & \eta_i + \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) + \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \geq \hat{\theta}_i(x_i) & \forall i, x_i \\
& \eta^k_{\alpha^k_{\mathcal{F}}} + \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) - \sum_{i\in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \geq 0 & \forall k, \alpha^k_{\mathcal{F}}, \boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \\
& \eta^k_\alpha - \delta^k_\alpha(x_i) - \nu^k_\alpha(\boldsymbol{x}_\alpha) \geq \hat{\theta}_\alpha(\boldsymbol{x}_\alpha) & \forall k, i\in\mathcal{V}^k, \alpha\in N^k_{\mathcal{F}}(i), \boldsymbol{x}_\alpha \\
& \eta^k_\alpha - \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) - \nu^k_\alpha(\boldsymbol{x}_\alpha) \geq \hat{\theta}_\alpha(\boldsymbol{x}_\alpha) & \forall k, \alpha^k_{\mathcal{F}}, \alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}}), \boldsymbol{x}_\alpha \\
& \sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha) \geq 0 & \forall \alpha, \boldsymbol{x}_\alpha
\end{aligned}
\tag{44}
$$

Note we reverse the sign of δ, λ and ν. By substituting η back into the dual objective function, we have the expression in Equ 42. Note that when $\sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \nu^k_\alpha(\boldsymbol{x}_\alpha) > 0$, we can always construct $\nu\prime$ satisfying $\sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} {\nu'_\alpha}^k(\boldsymbol{x}_\alpha) = 0$ with which the objective value is not increased. Thus we can replace the last inequality with equality. □

8.1.5 Proof of Claim 5

Claim. *When ν is fixing, the objective function of Equ 42 is lower bounded by*

$$
\begin{aligned}
& \min_\lambda \sum_k \left\{ \sum_{i\in\mathcal{V}^k} \max_{x_i} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) + \sum_{\alpha^k_{\mathcal{F}}} \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i\in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right) \right\} \\
& \text{where} \quad \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \max_{\boldsymbol{x}_{\alpha\setminus\alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) \right) \qquad \gamma^k_\alpha(x_i) = \max_{\boldsymbol{x}_\alpha\setminus x_i} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) \right) \\
& \qquad\qquad \hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \sum_{\alpha\in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \qquad \hat{\theta}_i(x_i) = \theta_i(x_i) + \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \gamma^k_\alpha(x_i)
\end{aligned}
\tag{45}
$$

Proof. **Derivation for node terms**

$$
\begin{aligned}
L_i =& \max_{x_i} \left(\theta_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) - \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) \right) + \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \max_{x_i} \left(\max_{\boldsymbol{x}_\alpha\setminus x_i} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) \right) + \delta^k_\alpha(x_i) \right) \\
=& \max_{x_i} \left(\theta_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) - \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \delta^k_\alpha(x_i) \right) + \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \max_{x_i} \left(\gamma^k_\alpha(x_i) + \delta^k_\alpha(x_i) \right) \\
\geq& \max_{x_i} \left(\theta_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) + \sum_{\alpha\in N^k_{\mathcal{F}}(i)} \gamma^k_\alpha(x_i) \right) \\
=& \max_{x_i} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}}\in N^k(i)} \lambda^k_{i\to\alpha^k_{\mathcal{F}}}(x_i) \right)
\end{aligned}
\tag{46}
$$

Derivation for free var set terms

$$\begin{aligned}
L_{\alpha^k_{\mathcal{F}}} &= \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) + \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\max_{\boldsymbol{x}_{\alpha \setminus \alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha) + \nu^k_\alpha(\boldsymbol{x}_\alpha) \right) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \right) \\
&= \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(- \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) + \sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \right) \\
&\geq \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\sum_{\alpha \in N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})} \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) \\
&= \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right)
\end{aligned} \tag{47}$$

□

8.1.6 Proof of Claim 6

Claim. *The message passing rule for λ is exactly the same as convex BP rule:*

$$\begin{aligned}
&\forall \alpha^k_{\mathcal{F}}, i \in N^k(\alpha^k_{\mathcal{F}}), x_i \quad \lambda_{i \to \alpha^k_{\mathcal{F}}}(x_i) = \frac{1}{1 + |N^k(i)|} \left(\hat{\theta}_i(x_i) + \sum_{\beta^k_{\mathcal{F}} \in N^k(i)} \mu_{\beta^k_{\mathcal{F}} \to i}(x_i) \right) - \mu_{\alpha^k_{\mathcal{F}} \to i}(x_i) \\
&\text{where} \qquad \mu_{\alpha^k_{\mathcal{F}} \to i}(x_i) = \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \setminus x_i} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{j \in N^k(\alpha^k_{\mathcal{F}}) \setminus i} \lambda_{j \to \alpha^k_{\mathcal{F}}}(x_j) \right)
\end{aligned} \tag{48}$$

The block coordinates rule for variables ν is

$$\nu^k_\alpha(\boldsymbol{x}_\alpha) = \frac{1}{|N_{\mathcal{P}}(\alpha)|} \sum_{j : \alpha \cap \mathcal{V}^j \neq \emptyset} \delta^j_\alpha(\boldsymbol{x}_{\alpha \cap \mathcal{V}^j}) - \delta^k_\alpha(\boldsymbol{x}_{\alpha \cap \mathcal{V}^k}) \tag{49}$$

where $|N_{\mathcal{P}}(\alpha)|$ is the number of sub-program in which α is involved. For arbitrary configuration of λ, variables δ can be decoded as

$$\delta^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) = \frac{1}{1 + |N^k_{\mathcal{F}}(\alpha^k_{\mathcal{F}})|} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{j \in N^k(\alpha^k_{\mathcal{F}})} \lambda_{j \to \alpha^k_{\mathcal{F}}}(x_j) \right) - \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) \tag{50}$$

and

$$\delta^k_\alpha(x_i) = \frac{1}{1 + |N^k_{\mathcal{F}}(i)|} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) - \gamma^k_\alpha(x_i) \tag{51}$$

Proof. From Equ 45, the sum of terms involving node i is

$$\begin{aligned}
&\max_{x_i} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) + \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{i \in N^k(\alpha^k_{\mathcal{F}})} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) \\
&= \max_{x_i} \left(\hat{\theta}_i(x_i) - \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) + \sum_{\alpha^k_{\mathcal{F}} \in N^k(i)} \max_{x_i} \left(\max_{\boldsymbol{x}_{\alpha^k_{\mathcal{F}}} \setminus x_i} \left(\hat{\theta}_{\alpha^k_{\mathcal{F}}}(\boldsymbol{x}_{\alpha^k_{\mathcal{F}}}) + \sum_{j \in N^k(\alpha^k_{\mathcal{F}}) \setminus i} \lambda^k_{j \to \alpha^k_{\mathcal{F}}}(x_j) \right) + \lambda^k_{i \to \alpha^k_{\mathcal{F}}}(x_i) \right) \\
&\geq \max_{x_i} \left(\hat{\theta}_i(x_i) + \sum_{\beta^k_{\mathcal{F}} \in N^k(i)} \mu_{\beta^k_{\mathcal{F}} \to i}(x_i) \right)
\end{aligned} \tag{52}$$

It is a lower bound achievable with Equ 48, which gives the block coordinate descent rule over λ.

From Equ 42, the sum of terms involving $\nu_\alpha^k(\boldsymbol{x}_\alpha)$ is

$$\sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset} \max_{\boldsymbol{x}_\alpha}\left(\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu_\alpha^k(\boldsymbol{x}_\alpha)+\delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})\right) \geq \max_{\boldsymbol{x}_\alpha}\left(\theta_\alpha(\boldsymbol{x}_\alpha)+\sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset}\delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})\right) \tag{53}$$

The lower bound can be achived by setting
$\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu_\alpha^k(\boldsymbol{x}_\alpha)+\delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = \frac{1}{|N_{\mathcal{P}}(\alpha)|}\left(\theta_\alpha(\boldsymbol{x}_\alpha)+\sum_{k:\alpha\cap\mathcal{V}^k\neq\emptyset}\delta_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})\right)$, which is equivalent to the rule in Equ 49.

By evaluate δ with Equ 50 and 51, the lower bound in Equ 45 can be achived, which gives the coordinate ascent rule for δ □

8.2 Another Expression of Sum-Product

The algorithm can be identically expressed as the following. In the implementation, we use the computation procedure shown in this expression.

Algorithm 5 Inference

Input: $\theta_i(x_i)$, $\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)=\frac{\theta_\alpha(\boldsymbol{x}_\alpha)}{|N_{\mathcal{P}}(\alpha)|}$
while Until convergency **do**
  for all k **do**
    $\forall\alpha_{\mathcal{F}}^k, \alpha\in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k), \boldsymbol{x}_{\alpha_{\mathcal{F}}^k}$, $\gamma_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_{\alpha\setminus\alpha_{\mathcal{F}}^k}} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu_\alpha^k(\boldsymbol{x}_\alpha)}{\epsilon c_\alpha}\right)$
    $\forall i\in\mathcal{V}^k, \alpha\in N_{\mathcal{F}(i)}^k, x_i$, $\gamma_\alpha^k(x_i) = \epsilon c_\alpha \ln \sum_{\boldsymbol{x}_\alpha\setminus x_i} \exp\left(\frac{\hat{\theta}_\alpha(\boldsymbol{x}_\alpha)+\nu_\alpha^k(\boldsymbol{x}_\alpha)}{\epsilon c_\alpha}\right)$
    $\forall\alpha_{\mathcal{F}}^k, \boldsymbol{x}_{\alpha_{\mathcal{F}}^k}$, $\hat{\theta}_{\alpha_{\mathcal{F}}^k}(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k}) = \sum_{\alpha\in N_{\mathcal{F}}^k(\alpha_{\mathcal{F}}^k)} \gamma_\alpha^k(\boldsymbol{x}_{\alpha_{\mathcal{F}}^k})$
    $\forall i\in\mathcal{V}^k, x_i$, $\hat{\theta}_i(x_i) = \theta_i(x_i) + \sum_{\alpha\in N_{\mathcal{F}}^k(i)} \gamma_\alpha^k(x_i)$
  end for
  for all k **do**
    δ_α^k = Sub-Inference($\hat{\theta}_i$, $\hat{\theta}_{\alpha_{\mathcal{F}}^k}$, γ_α^k)
  end for
  for all k **do**
    $\nu_\alpha^k(\boldsymbol{x}_\alpha) = \frac{1}{|N_{\mathcal{P}}(\alpha)|}\sum_{j:\alpha\cap\mathcal{V}^j\neq\emptyset}\delta_\alpha^j(\boldsymbol{x}_{\alpha\cap\mathcal{V}^j}) - \delta_\alpha^k(\boldsymbol{x}_{\alpha\cap\mathcal{V}^k})$
  end for
end while

Algorithm 6 Sub-Inference

Input: $\gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}})$, $\gamma^k_\alpha(x_i)$, $\hat{\theta}_{\alpha^k_\mathcal{F}}(\boldsymbol{x}_{\alpha^k_\mathcal{F}})$, $\hat{\theta}_i(x_i)$
for all $t \leq$ max num of inner iter **do**
  for all $i \in \mathcal{V}^k$ **do**
    $\forall \alpha^k_\mathcal{F} \in N^k(i), x_i, \quad \lambda_{i \to \alpha^k_\mathcal{F}}(x_i) = \frac{\hat{c}_{\alpha^k_\mathcal{F}}}{\hat{c}_i}\left(\hat{\theta}_i(x_i) + \sum_{\beta^k_\mathcal{F} \in N^k(i)} \mu_{\beta^k_\mathcal{F} \to i}(x_i)\right) - \mu_{\alpha^k_\mathcal{F} \to i}(x_i)$
    $\forall \alpha^k_\mathcal{F} \in N^k(i), x_i, \quad \mu_{\alpha^k_\mathcal{F} \to i}(x_i) = \epsilon \hat{c}_{\alpha^k_\mathcal{F}} \ln \sum_{\boldsymbol{x}_{\alpha^k_\mathcal{F}} \setminus x_i} \exp\left(\frac{\hat{\theta}_{\alpha^k_\mathcal{F}}(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) + \sum_{j \in N^k(\alpha^k_\mathcal{F}) \setminus i} \lambda_{j \to \alpha^k_\mathcal{F}}(x_j)}{\epsilon \hat{c}_{\alpha^k_\mathcal{F}}}\right)$
  end for
  for all $\alpha : \alpha \cap \mathcal{V}^k \neq \emptyset$ **do**
    if $\alpha \cap \mathcal{V}^k$ is a single node **then**
      $\delta^k_\alpha(x_i) = \frac{c_\alpha}{\hat{c}_i}\left(\hat{\theta}_i(x_i) + \sum_{\alpha^k_\mathcal{F} \in N^k(i)} \mu_{\alpha^k_\mathcal{F} \to i}(x_i)\right) - \gamma^k_\alpha(x_i)$
    else
      $\delta^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) = \frac{c_\alpha}{\hat{c}_{\alpha^k_\mathcal{F}}}\left(\hat{\theta}_{\alpha^k_\mathcal{F}}(\boldsymbol{x}_{\alpha^k_\mathcal{F}}) + \sum_{j \in N^k(\alpha^k_\mathcal{F})} \lambda_{j \to \alpha^k_\mathcal{F}}(x_j)\right) - \gamma^k_\alpha(\boldsymbol{x}_{\alpha^k_\mathcal{F}})$
    end if
  end for
end for
Return δ^k_α