Feedback Loop between High Level Semantics and Low Level Vision - Supplementary Material

Varun K. Nagaraja Vlad I. Morariu Larry S. Davis
{varun,morariu,lsd}@umiacs.umd.edu

University of Maryland, College Park, MD, USA.

Before deriving the score functions, we first formulate the MAP inference problem in binary Markov networks as an Integer Linear Program (ILP) following the work of Globerson and Jaakkola [1]. The integer variables in the ILP are then relaxed to continuous values giving us a relaxed linear program. We then obtain the dual of this relaxed linear program and show a block co-ordinate descent strategy that can be used to solve the dual through a "message passing algorithm". However we do not solve the the inference problem in the dual space. We only use the message update equations for the dual variables to obtain our score functions.

Sontag et al. [2] use these message update equations to rank clusters of variables in their cluster pursuit algorithm which incrementally adds clusters of variables to the objective function and solves the MAP problem in the dual space. They rank the clusters using a score function that measures the decrease in the dual value of the objective function when a cluster is added. We also derive our score functions similar to their approach by using the message update equations.

While the derivations in [2,3] are provided for pairwise graphical models, we derive them for general networks of any order.

1 Linear Programming Relaxation of the MAP problem

Let $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ be a set of binary variables and $\mathcal{C} = \{c : c \subset (1, 2, \ldots n)\}$ be a set of clusters. Consider a function $\Phi(\boldsymbol{x}; \boldsymbol{\theta})$ defined as a sum of the functions $\theta_c(x_c)$ defined over the clusters[1] of variables. The goal of Maximum A Posteriori assignment (MAP) is to find an assignment that maximizes the function $\Phi(\boldsymbol{x}; \boldsymbol{\theta})$.

$$\arg\max_{\boldsymbol{x}} \Phi(\boldsymbol{x}; \boldsymbol{\theta}) = \arg\max_{\boldsymbol{x}} \sum_{c \in \mathcal{C}} \theta_c(x_c) \tag{1}$$

Let $\mathcal{S} = \{c \cap c' : c, c' \in \mathcal{C}, c \cap c' \neq \emptyset\}$ be the set of intersections between clusters and $\mathcal{S}(c) = \{s \in \mathcal{S} : s \subseteq c\}$ be the set of overlap sets for cluster c. The above problem can be reformulated as an integer program by introducing indicator variables $\mu_c(x_c)$ for each cluster, $\mu_s(x_s)$ for each intersection set between clusters

[1] The analysis does not require an underlying graph and hence the clusters need not correspond to cliques.

and $\mu_i(x_i)$ for each variable.

$$\begin{aligned}
\underset{\boldsymbol{\mu}}{\text{maximize}} \quad & \sum_{c \in \mathcal{C}} \sum_{x_c} \mu_c(x_c)\theta_c(x_c) && (2)\\
\text{subject to} \quad & \mu_c(x_c) \in \{0,1\} \quad \forall c \in \mathcal{C} && (3)\\
& \sum_{x_i} \mu_i(x_i) = 1 \quad \forall i \in \{1,\dots,n\} && (4)\\
& \sum_{x_{s \setminus i}} \mu_s(x_s) = \mu_i(x_i) \quad \forall s \in \mathcal{S}, i \in s && (5)\\
& \sum_{x_{c \setminus s}} \mu_c(x_c) = \mu_s(x_s) \quad \forall c \in \mathcal{C}, s \in \mathcal{S}(c) && (6)
\end{aligned}$$

The constraint in Equation (6) enforces that the cluster indicator variables must be consistent with the intersection set indicator variable and the constraint in Equation (5) enforces the consistency of an individual variable with all the intersection sets that it is part of. The set of constraints on $\boldsymbol{\mu}$ denoted as $\mathcal{M}_L(\mathcal{C})$ is known as the marginal polytope. This problem is completely equivalent to the original problem 1 and is hence as hard as the original problem. In many cases, this is NP-Hard and hence we obtain a linear programming relaxation by allowing the indicator variables to take on non-integer values i.e. replace the constraints as $\mu_c(x_c) \in [0,1]$. The optimum of the relaxed problem gives an upper bound on the MAP value.

We will now find the dual problem of the relaxed linear program. Let $\lambda_{cs}(x_s)$ and $\lambda_{si}(x_i)$ be the dual variables corresponding to each of the constraints in Equation (6) and Equation (5) respectively. The constraint in Equation (4) will be kept implicit and used to simplify the Lagrangian later.

The Lagrangian is given by

$$\begin{aligned}
L(\boldsymbol{\mu},\boldsymbol{\lambda}) = & \sum_{c \in \mathcal{C}} \sum_{x_c} \mu_c(x_c)\theta_c(x_c) + \sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}(c)} \sum_{x_s} \lambda_{cs}(x_s) \left[\mu_s(x_s) - \sum_{x_{c \setminus s}} \mu_c(x_c) \right] \\
& + \sum_{s \in \mathcal{S}} \sum_{i \in s} \sum_{x_i} \lambda_{si}(x_i) \left[\mu_i(x_i) - \sum_{x_{s \setminus i}} \mu_s(x_s) \right] \qquad (7)
\end{aligned}$$

After rearranging the terms to group by common indicator variables, we get

$$\begin{aligned}
L(\boldsymbol{\mu},\boldsymbol{\lambda}) = & \sum_{c \in \mathcal{C}} \sum_{x_c} \mu_c(x_c) \left[\theta_c(x_c) - \sum_{s \in \mathcal{S}(c)} \lambda_{cs}(x_s) \right] \\
& + \sum_{s \in \mathcal{S}} \sum_{x_s} \mu_s(x_s) \left[\sum_{c : s \in \mathcal{S}(c)} \lambda_{cs}(x_s) - \sum_{i \in s} \lambda_{si}(x_i) \right] \\
& + \sum_{i} \sum_{x_i} \mu_i(x_i) \left[\sum_{s : i \in s} \lambda_{si}(x_i) \right] \qquad (8)
\end{aligned}$$

We can now analytically maximize with respect to $\boldsymbol{\mu} \geq 0$ and the implicit constraint in Equation (4) to obtain the dual objective function,

$$\begin{aligned} J(\boldsymbol{\lambda}) &= \max_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \boldsymbol{\lambda}) \\ &= \sum_{c \in \mathcal{C}} \max_{x_c} \left[\theta_c(x_c) - \sum_{s \in \mathcal{S}(c)} \lambda_{cs}(x_s) \right] \\ &\quad + \sum_{s \in \mathcal{S}} \max_{x_s} \left[\sum_{c : s \in \mathcal{S}(c)} \lambda_{cs}(x_s) - \sum_{i \in s} \lambda_{si}(x_i) \right] + \sum_i \max_{x_i} \left[\sum_{s : i \in s} \lambda_{si}(x_i) \right] \end{aligned} \quad (9)$$

The unconstrained dual program is now just

$$\underset{\boldsymbol{\lambda}}{\text{minimize}} \quad J(\boldsymbol{\lambda}) \quad (10)$$

The above dual formulation is a simple extension of the technique adopted by D. Sontag [4] where they derive the dual of the LP relaxation for pairwise potentials. Another dual formulation, with constraints, can also be obtained by following the method of Globerson and Jaakkola [1].

2 Block Coordinate Descent in the Dual

A block coordinate descent strategy can be used to minimize the dual objective. At every iteration, the dual variables $\lambda_{cs}(x_s)$ are updated for one cluster while the rest are kept fixed. Similarly the dual variables $\lambda_{si}(x_i)$ are updated for one intersection set at a time while the rest are kept fixed. The update messages for the dual variables are given below.

From a cluster to one of its intersection sets:

$$\begin{aligned} \lambda_{cs}(x_s) &= -\lambda_s^{-c}(x_s) - \sum_{i \in s} \lambda_{si}(x_i) \\ &\quad + \frac{1}{|\mathcal{S}(c)|} \max_{x_{c \setminus s}} \left[\theta_c(x_c) + \sum_{\hat{s} \in \mathcal{S}(c)} \lambda_{\hat{s}}^{-c}(x_{\hat{s}}) - \sum_{\hat{s} \in \mathcal{S}(c)} \sum_{i \in \hat{s}} \lambda_{\hat{s}i}(x_i) \right] \end{aligned} \quad (11)$$

where

$$\lambda_s^{-c}(x_s) = \sum_{\hat{c} \neq c : s \in \mathcal{S}(\hat{c})} \lambda_{\hat{c}s}(x_s) \quad (12)$$

From an intersection set to one of its variables:

$$\lambda_{si}(x_i) = -\lambda_i^{-s}(x_i) + \frac{1}{|s|} \max_{x_{s \setminus i}} \left[\sum_{c : s \in \mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{\hat{i} \in s} \lambda_{\hat{i}}^{-s}(x_{\hat{i}}) \right] \quad (13)$$

where

$$\lambda_i^{-s}(x_i) = \sum_{\hat{s} \neq s : i \in \hat{s}} \lambda_{\hat{s}i}(x_i) \quad (14)$$

The derivation of the update messages can be found in Section 5.

3 Upper Bound Score - Proof of Proposition 1

Proposition 1 (Upper Bound Score). *An upper bound on the change in the MAP value after adding a cluster is given by*

$$\varDelta\varPhi \leq \frac{1}{|s|}\sum_{i\in s}\max_{x_i}\left(\max_{x_{cur\setminus i}}\varPhi_{cur}(x_{cur}) + \max_{x_{new\setminus i}}\varPhi_{new}(x_{new})\right) - \max\varPhi_{cur}(x_{cur}) \tag{15}$$

where s is the set of nodes in the intersection of the sets x_{cur} and x_{new}.

Proof. In the block coordinate descent algorithm, during each iteration of the minimization procedure, the dual variables $\lambda_{cs}(x_s)$ are updated for one cluster c and all its intersection sets $s \in \mathcal{S}(c)$ while the rest are kept fixed. Similarly the dual variables $\lambda_{si}(x_i)$ are updated for one intersection set s and all the variables in this set ($i \in s$) while the rest are kept fixed.

We calculate the scores for one cluster at a time while setting the dual variables for other clusters to zero. Let $\theta_f(x_f)$ be the cluster of potential functions of the *current* network and $\theta_g(x_g)$ be the cluster of potential functions of the *new* cluster. We start with all the dual variables set to zero. Since we consider only two clusters f and g, the number of intersection sets is just one i.e. $|\mathcal{S}(f)| = |\mathcal{S}(g)| = 1$. The first update (Equation 11) is performed to the dual variable $\lambda_{fs}(x_s)$ while setting the rest of the dual variables to zero.

$$\lambda_{fs}(x_s) = \max_{x_{f\setminus s}}[\theta_f(x_f)] \tag{16}$$

This is followed by an update to the dual variable $\lambda_{gs}(x_s)$ given by

$$\lambda_{gs}(x_s) = -\lambda_{fs}(x_s) + \max_{x_{g\setminus s}}[\theta_g(x_g) + \lambda_{fs}(x_s)] = \max_{x_{g\setminus s}}\theta_g(x_g) \tag{17}$$

Finally we update (Equation 13) the dual variables $\lambda_{si}(x_i)$

$$\lambda_{si}(x_i) = \frac{1}{|s|}\max_{x_{s\setminus i}}[\lambda_{fs}(x_s) + \lambda_{gs}(x_s)] \tag{18}$$

We now measure the value of the dual objective function before and after updating the dual variables. The dual objective function (Equation 9) in our case is given by

$$\begin{aligned} J = &\max_{x_f}[\theta_f(x_f) - \lambda_{fs}(x_s)] + \max_{x_g}[\theta_g(x_g) - \lambda_{gs}(x_s)] \\ &+ \max_{x_s}\left[\lambda_{fs}(x_s) + \lambda_{gs}(x_s) - \sum_{i\in s}\lambda_{si}(x_i)\right] \\ &+ \sum_{i\in s}\max_{x_i}[\lambda_{si}(x_i)] \end{aligned} \tag{19}$$

When we initialize the dual variables to zero, the value of the dual objective function is

$$J^{(0)} = \max_{x_f} \theta_f(x_f) + \max_{x_g} \theta_g(x_g) \tag{20}$$

After performing one update for the dual variables $\lambda_{fs}(x_s)$ (Equation 16) and $\lambda_{gs}(x_s)$ (Equation 17), we can see that

$$\max_{x_f} [\theta_f(x_f) - \lambda_{fs}(x_s)] = \max_{x_f} \left[\theta_f(x_f) - \max_{x_{f\setminus s}}[\theta_f(x_f)]\right] \leq 0 \tag{21}$$

$$\max_{x_g} [\theta_g(x_g) - \lambda_{gs}(x_s)] = \max_{x_g} \left[\theta_g(x_g) - \max_{x_{g\setminus s}}[\theta_g(x_g)]\right] \leq 0 \tag{22}$$

Also substituting for $\lambda_{si}(x_i)$ from Equation (18) gives us

$$\begin{aligned} &\max_{x_s} \left[\lambda_{fs}(x_s) + \lambda_{gs}(x_s) - \sum_{i \in s} \lambda_{si}(x_i)\right] \\ &= \max_{x_s} \left[\lambda_{fs}(x_s) + \lambda_{gs}(x_s) - \sum_{i \in s} \frac{1}{|s|} \max_{x_{s\setminus i}} [\lambda_{fs}(x_s) + \lambda_{gs}(x_s)]\right] \leq 0 \end{aligned} \tag{23}$$

Hence the dual value after performing one update of the dual variables is

$$J^{(1)} \leq \sum_{i \in s} \max_{x_i} [\lambda_{si}(x_i)] \tag{24}$$

To avoid performing costly max-marginalization over the intersection set s to calculate $\lambda_{fs}(x_s)$ and $\lambda_{gs}(x_s)$, we can approximate $\lambda_{si}(x_i)$ as follows

$$\lambda_{si}(x_i) = \frac{1}{|s|} \max_{x_{s\setminus i}} [\lambda_{fs}(x_s) + \lambda_{gs}(x_s)] \tag{25}$$

$$\leq \frac{1}{|s|} \left(\max_{x_{s\setminus i}} \lambda_{fs}(x_s) + \max_{x_{s\setminus i}} \lambda_{gs}(x_s)\right) \tag{26}$$

$$= \frac{1}{|s|} \left(\max_{x_{f\setminus i}} \theta_f(x_f) + \max_{x_{g\setminus i}} \theta_g(x_g)\right) \tag{27}$$

We still need to perform max-marginalization, but only over one variable at a time. This gives us a new upper bound on the dual value

$$J^{(1)} \leq \frac{1}{|s|} \sum_{i \in s} \max_{x_i} \left[\max_{x_{f\setminus i}} \theta_f(x_f) + \max_{x_{g\setminus i}} \theta_g(x_g)\right] \tag{28}$$

Since the dual value is an upper bound on the primal MAP value, we have

$$\max_{x} [\theta_f(x_f) + \theta_g(x_g)] \leq \frac{1}{|s|} \sum_{i \in s} \max_{x_i} \left[\max_{x_{f\setminus i}} \theta_f(x_f) + \max_{x_{g\setminus i}} \theta_g(x_g)\right] \tag{29}$$

Substituting for cluster $\theta_f(x_f)$ as $\Phi_{\text{cur}}(x_{\text{cur}})$ and $\theta_g(x_g)$ as $\Phi_{\text{new}}(x_{\text{new}})$ we can write an upper bound for the change in the primal MAP value after adding a cluster as

$$\Delta\Phi \leq \frac{1}{|s|}\sum_{i\in s}\max_{x_i}\left(\max_{x_{\text{cur}\setminus i}}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i}}\Phi_{\text{new}}(x_{\text{new}})\right) - \max\Phi_{\text{cur}}(x_{\text{cur}}) \tag{30}$$

4 Blind Score - Proof of Proposition 2

Proposition 2 (Blind Score). *A lower bound to the upper bound score (15) is given by*

$$score(g)_{upper} \geq score(g)_{blind} \tag{31}$$

$$= \frac{-1}{|s|}\sum_{i\in s}\left|\max_{x_i=0,x_{cur\setminus i}}\Phi_{cur}(x_{cur}) - \max_{x_i=1,x_{cur\setminus i}}\Phi_{cur}(x_{cur})\right| \tag{32}$$

where s is the set of nodes in the intersection of the sets x_{cur} and x_{new}.

Proof.

$$score(g)_{upper}$$

$$= \frac{1}{|s|}\sum_{i\in s}\max_{x_i}\left(\max_{x_{\text{cur}\setminus i}}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i}}\Phi_{\text{new}}(x_{\text{new}})\right) - \max\Phi_{\text{cur}}(x_{\text{cur}}) \tag{33}$$

$$= \frac{1}{|s|}\sum_{i\in s}\max\left\{\begin{array}{l}\left(\max_{x_{\text{cur}\setminus i},x_i=0}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i},x_i=0}\Phi_{\text{new}}(x_{\text{new}})\right),\\ \left(\max_{x_{\text{cur}\setminus i},x_i=1}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i},x_i=1}\Phi_{\text{new}}(x_{\text{new}})\right)\end{array}\right\}$$

$$- \max\Phi_{\text{cur}}(x_{\text{cur}}) \tag{34}$$

$$= \frac{1}{|s|}\sum_{i\in s}\delta_i \tag{35}$$

where

$$\delta_i = \max\left\{\left(\max_{x_{\text{cur}\setminus i},x_i=1}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i},x_i=1}\Phi_{\text{new}}(x_{\text{new}}) - \max\Phi_{\text{cur}}(x_{\text{cur}})\right),\right.$$
$$\left.\left(\max_{x_{\text{cur}\setminus i},x_i=0}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i},x_i=0}\Phi_{\text{new}}(x_{\text{new}}) - \max\Phi_{\text{cur}}(x_{\text{cur}})\right)\right\} \tag{36}$$

Let us assume that the assignment to some $x_i = 1$ in $\max\Phi_{\text{cur}}(x_{\text{cur}})$. Then δ_i becomes

$$\delta_i = \max\left\{\max_{x_{\text{new}\setminus i},x_i=1}\Phi_{\text{new}}(x_{\text{new}}),\right.$$
$$\left.\left(\max_{x_{\text{cur}\setminus i},x_i=0}\Phi_{\text{cur}}(x_{\text{cur}}) + \max_{x_{\text{new}\setminus i},x_i=0}\Phi_{\text{new}}(x_{\text{new}}) - \max_{x_{\text{cur}\setminus i},x_i=1}\Phi_{\text{cur}}(x_{\text{cur}})\right)\right\} \tag{37}$$

We now assume that $\max_{x_{\text{new}\setminus i}, x_i=0} \Phi_{\text{new}}(x_{\text{new}}) \geq 0$, which can be enforced by adding a positive offset to $\Phi_{\text{new}}(x_{\text{new}})$. A lower bound for δ_i is then

$$\delta_i \geq \left(\max_{x_{\text{cur}\setminus i}, x_i=0} \Phi_{\text{cur}}(x_{\text{cur}}) - \max_{x_{\text{cur}\setminus i}, x_i=1} \Phi_{\text{cur}}(x_{\text{cur}}) \right) \tag{38}$$

Since the maximizing assignment to $\Phi_{\text{cur}}(x_{\text{cur}})$ had $x_i = 1$, any other assignment with $x_i = 0$ must be less than the maxima. Hence,

$$\delta_i \geq - \left| \max_{x_{\text{cur}\setminus i}, x_i=0} \Phi_{\text{cur}}(x_{\text{cur}}) - \max_{x_{\text{cur}\setminus i}, x_i=1} \Phi_{\text{cur}}(x_{\text{cur}}) \right| \tag{39}$$

A similar argument can be made if the assignment to an $x_i = 0$. Hence we can put all the δ_i together to obtain a lower bound on the upper bound score

$$score(g)_{upper} = \frac{1}{|s|} \sum_{i \in s} \delta_i \tag{40}$$

$$\geq \frac{-1}{|s|} \sum_{i \in s} \left| \max_{x_{\text{cur}\setminus i}, x_i=0} \Phi_{\text{cur}}(x_{\text{cur}}) - \max_{x_{\text{cur}\setminus i}, x_i=1} \Phi_{\text{cur}}(x_{\text{cur}}) \right| \tag{41}$$

5 Message Update Equations

Theorem 1. *The message update in Equation (11) for the dual variable $\lambda_{cs}(x_s)$ corresponds to block co-ordinate descent on the dual objective $J(\boldsymbol{\lambda})$.*

Proof. The proof follows from the ideas in the derivation of the optimality of the MPLP update from [3]. It shows that the value of the dual objective function reaches the minima in the variable λ_{cs} after performing a single update to it.

Consider fixing all $\lambda_{cs}(x_s)$ except for one cluster c. The part of the objective function that is dependent on the free variables is given by

$$\bar{J}(\boldsymbol{\lambda}) = \max_{x_c} \left[\theta_c(x_c) - \sum_{s \in \mathcal{S}(c)} \lambda_{cs}(x_s) \right] + \sum_{s \in \mathcal{S}(c)} \max_{x_s} \left[\sum_{c : s \in \mathcal{S}(c)} \lambda_{cs}(x_s) - \sum_{i \in s} \lambda_{si}(x_i) \right] \tag{42}$$

Let

$$\lambda_s^{-c}(x_s) = \sum_{\hat{c} \neq c : s \in \mathcal{S}(\hat{c})} \lambda_{\hat{c}s}(x_s) \tag{43}$$

then $\bar{J}(\boldsymbol{\lambda})$ can be rewritten as

$$\begin{aligned}\bar{J}(\boldsymbol{\lambda}) &= \max_{x_c}\left[\theta_c(x_c) - \sum_{s\in\mathcal{S}(c)} \lambda_{cs}(x_s)\right] \\ &\quad + \sum_{s\in\mathcal{S}(c)} \max_{x_s}\left[\lambda_{cs}(x_s) + \lambda_s^{-c}(x_s) - \sum_{i\in s}\lambda_{si}(x_i)\right] &(44)\\ &= A_c(x_c) + \sum_{s\in\mathcal{S}(c)} A_s(x_s) &(45)\end{aligned}$$

The lower bound on $\bar{J}(\boldsymbol{\lambda})$ is given by

$$\begin{aligned}\bar{J}(\boldsymbol{\lambda}) &\geq \max_{x_c}\left(\left[\theta_c(x_c) - \sum_{s\in\mathcal{S}(c)} \lambda_{cs}(x_s)\right]\right. \\ &\quad \left. + \sum_{s\in\mathcal{S}(c)} \left[\lambda_{cs}(x_s) + \lambda_s^{-c}(x_s) - \sum_{i\in s}\lambda_{si}(x_i)\right]\right) &(46)\\ &= \max_{x_c}\left(\theta_c(x_c) + \sum_{s\in\mathcal{S}(c)} \lambda_s^{-c}(x_s) - \sum_{s\in\mathcal{S}(c)}\sum_{i\in s}\lambda_{si}(x_i)\right) = B &(47)\end{aligned}$$

If we apply the update messages in Equation (11) to $A_c(x_c)$, we get

$$\begin{aligned}A_c(x_c) &= \max_{x_c}\left[\theta_c(x_c) - \sum_{s\in\mathcal{S}(c)} \lambda_{cs}(x_s)\right] &(48)\\ &= \max_{x_c}\left[\theta_c(x_c) + \sum_{s\in\mathcal{S}(c)} \lambda_s^{-c}(x_s) + \sum_{s\in\mathcal{S}(c)}\sum_{i\in s}\lambda_{si}(x_i)\right. \\ &\quad \left. -\frac{1}{|\mathcal{S}(c)|}\sum_{s\in\mathcal{S}(c)} \max_{x_{c\setminus s}}\left[\theta_c(x_c) + \sum_{\hat{s}\in\mathcal{S}(c)} \lambda_{\hat{s}}^{-c}(x_{\hat{s}}) - \sum_{\hat{s}\in\mathcal{S}(c)}\sum_{i\in \hat{s}}\lambda_{\hat{s}i}(x_i)\right]\right] &(49)\\ &\leq \max_{x_c}\left[\theta_c(x_c) + \sum_{s\in\mathcal{S}(c)} \lambda_s^{-c}(x_s) + \sum_{s\in\mathcal{S}(c)}\sum_{i\in s}\lambda_{si}(x_i)\right] \\ &\quad -\frac{1}{|\mathcal{S}(c)|}\sum_{s\in\mathcal{S}(c)} \max_{x_c}\left[\theta_c(x_c) + \sum_{\hat{s}\in\mathcal{S}(c)} \lambda_{\hat{s}}^{-c}(x_{\hat{s}}) - \sum_{\hat{s}\in\mathcal{S}(c)}\sum_{i\in \hat{s}}\lambda_{\hat{s}i}(x_i)\right] &(50)\\ &= 0 &(51)\end{aligned}$$

Similarly by applying the update to $A_x(x_s)$, we get

$$A_s(x_s) = \max_{x_s}\left[\lambda_s^{-c}(x_s) - \sum_{i\in s}\lambda_{si}(x_i) + \lambda_{cs}(x_s)\right] \tag{52}$$

$$= \max_{x_s}\Bigg[\lambda_s^{-c}(x_s) - \sum_{i\in s}\lambda_{si}(x_i) - \lambda_s^{-c}(x_s) - \sum_{i\in s}\lambda_{si}(x_i) + \frac{1}{|\mathcal{S}(c)|}\max_{x_{c\setminus s}}\left[\theta_c(x_c) + \sum_{\hat{s}\in\mathcal{S}(c)}\lambda_{\hat{s}}^{-c}(x_{\hat{s}}) - \sum_{\hat{s}\in\mathcal{S}(c)}\sum_{i\in\hat{s}}\lambda_{\hat{s}i}(x_i)\right]\Bigg] \tag{53}$$

$$= \frac{1}{|\mathcal{S}(c)|}\max_{x_c}\left[\theta_c(x_c) + \sum_{\hat{s}\in\mathcal{S}(c)}\lambda_{\hat{s}}^{-c}(x_{\hat{s}}) - \sum_{\hat{s}\in\mathcal{S}(c)}\sum_{i\in\hat{s}}\lambda_{\hat{s}i}(x_i)\right] \tag{54}$$

$$= \frac{B}{|\mathcal{S}(c)|} \tag{55}$$

Therefore

$$\bar{J}(\boldsymbol{\lambda}) = A_c(x_c) + \sum_{s\in\mathcal{S}(c)} A_s(x_s) \le B \tag{56}$$

whereas we earlier showed that B is the lower bound on $\bar{J}(\boldsymbol{\lambda})$. Hence $\bar{J}(\boldsymbol{\lambda}) = B$ which implies that the update equation does indeed minimize the dual objective in the coordinates $\lambda_{cs}(x_s)$.

Theorem 2. *The message update in Equation (13) for the dual variable $\lambda_{si}(x_i)$ corresponds to block co-ordinate descent on the dual objective $J(\boldsymbol{\lambda})$.*

Proof. Consider fixing all $\lambda_{si}(x_i)$ except for one intersection set s. The part of the objective function that is dependent on the free variables is given by

$$\bar{J}(\boldsymbol{\lambda}) = \max_{x_s}\left[\sum_{c:s\in\mathcal{S}(c)}\lambda_{cs}(x_s) - \sum_{i\in s}\lambda_{si}(x_i)\right] + \sum_{i\in s}\max_{x_i}\left[\sum_{s:i\in s}\lambda_{si}(x_i)\right] \tag{57}$$

$$= A_s(x_s) + \sum_{i\in s} A_i(x_i) \tag{58}$$

Let

$$\lambda_i^{-s}(x_i) = \sum_{\hat{s}\neq s:i\in\hat{s}}\lambda_{\hat{s}i}(x_i) \tag{59}$$

The lower bound on $\bar{J}(\boldsymbol{\lambda})$ is given by

$$\bar{J}(\boldsymbol{\lambda}) \ge \max_{x_s}\left[\sum_{c:s\in\mathcal{S}(c)}\lambda_{cs}(x_s) + \sum_{i\in s}\lambda_i^{-s}(x_i)\right] = B \tag{60}$$

When we apply the update in Equation (13) to $A_s(x_s)$ we get,

$$\begin{aligned} A_s(x_s) &= \max_{x_s} \left[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) - \sum_{i\in s} \lambda_{si}(x_i) \right] && (61) \\ &= \max_{x_s} \Bigg[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{i\in s} \lambda_i^{-s}(x_i) \\ &\qquad - \frac{1}{|s|} \sum_{i\in s} \max_{x_{s\setminus i}} \left[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{\hat{i}\in s} \lambda_{\hat{i}}^{-s}(x_{\hat{i}}) \right] \Bigg] && (62) \\ &\leq \max_{x_s} \left[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{i\in s} \lambda_i^{-s}(x_i) \right] \\ &\qquad - \frac{1}{|s|} \sum_{i\in s} \max_{x_s} \left[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{\hat{i}\in s} \lambda_{\hat{i}}^{-s}(x_{\hat{i}}) \right] && (63) \\ &= 0 && (64) \end{aligned}$$

Similarly by applying the update to $A_i(x_i)$, we get

$$\begin{aligned} A_i(x_i) &= \max_{x_i} \left[\lambda_{si}(x_i) + \lambda_i^{-s}(x_i) \right] && (65) \\ &= \max_{x_i} \left[\frac{1}{|s|} \max_{x_{s\setminus i}} \left[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{\hat{i}\in s} \lambda_{\hat{i}}^{-s}(x_{\hat{i}}) \right] \right] && (66) \\ &= \frac{1}{|s|} \max_{x_s} \left[\sum_{c:s\in\mathcal{S}(c)} \lambda_{cs}(x_s) + \sum_{\hat{i}\in s} \lambda_{\hat{i}}^{-s}(x_{\hat{i}}) \right] && (67) \\ &= \frac{B}{|s|} && (68) \end{aligned}$$

Therefore we get

$$\bar{J}(\boldsymbol{\lambda}) = A_s(x_s) + \sum_{i\in s} A_i(x_i) \leq B \qquad (69)$$

But we showed that $\bar{J}(\boldsymbol{\lambda}) \geq B$. Hence $\bar{J}(\boldsymbol{\lambda}) = B$ and the update equation minimizes the dual objective in the coordinates $\lambda_{si}(x_i)$.

References

1. Globerson, A., Jaakkola, T.: Fixing Max-Product: Convergent Message Passing Algorithms for MAP LP-Relaxations. In: NIPS. (2007)

2. Sontag, D., Meltzer, T., Globerson, A.: Tightening LP Relaxations for MAP using Message Passing. In: UAI. (2008)
3. Sontag, D., Globerson, A., Jaakkola, T.: Introduction to Dual Decomposition for Inference. (2011)
4. Sontag, D.A.: Approximate Inference in Graphical Models using LP Relaxations. PhD thesis, MIT (2010)