Partial Optimality of Dual Decomposition for MAP Inference in Pairwise MRFs - Supplementary Material

The following two lemmas are required for the proof of Theorem 1.

Lemma 3 *Let g_7, g_8 be the dual functions of the problems (7) and (8) (presented in (9) and (10)), respectively. For any value of the dual variables $\boldsymbol{u}$ the equality $g_7(\boldsymbol{u}) = g_8(\boldsymbol{u})$ holds.*

Proof:

$$\begin{aligned} g_8(\boldsymbol{u}) &= \inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1}, \dots, \boldsymbol{\mu}^m \in L_{\mathcal{T}_m}} \left\{ \sum_{j=1}^m (\boldsymbol{\theta}^j + \boldsymbol{u}^j)^\top \boldsymbol{\mu}^j \right\} \\ &= \inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1}} \{(\boldsymbol{\theta}^1 + \boldsymbol{u}^1)^\top \boldsymbol{\mu}^1\} + \dots + \inf_{\boldsymbol{\mu}^m \in L_{\mathcal{T}_m}} \{(\boldsymbol{\theta}^m + \boldsymbol{u}^m)^\top \boldsymbol{\mu}^m\} \\ &= \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1}} \{(\boldsymbol{\theta}^1 + \boldsymbol{u}^1)^\top \boldsymbol{\mu}^1\} + \dots + \inf_{\boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m}} \{(\boldsymbol{\theta}^m + \boldsymbol{u}^m)^\top \boldsymbol{\mu}^m\} \\ &= \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1}, \dots, \boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m}} \left\{ \sum_{j=1}^m (\boldsymbol{\theta}^j + \boldsymbol{u}^j)^\top \boldsymbol{\mu}^j \right\} = g_7(\boldsymbol{u}) \end{aligned}$$

In the third equation we used the known fact that for a tree-structured MRF the local consistency polytope coincides with the marginal polytope and that a corresponding objective for each subproblem is linear.

□

Lemma 4 *For any tree-structured MRF $\mathcal{T}$ with the corresponding set of valid assignments $\mathcal{X}_\mathcal{T} \subseteq \mathbb{R}^d$ in the standard overcomplete representation (as defined in (3)) and for any $\mathcal{I} \subseteq \{1, \dots, d\}$ the following equality holds*

$$\text{conv } \mathcal{X}_\mathcal{T} \cap \{\boldsymbol{\mu} \in \mathbb{R}^d : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\} = \text{conv } \{\boldsymbol{\mu} \in \mathcal{X}_\mathcal{T} : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\}, \tag{.13}$$

where $\boldsymbol{\mu}^* \in \mathbb{R}^d$ is a point, for which there exists an assignment $\bar{\boldsymbol{\mu}} \in \mathcal{X}_\mathcal{T}$ with $\bar{\boldsymbol{\mu}}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}$.

Proof: " $\subseteq$ " : Let $\boldsymbol{v} \in \text{conv } \mathcal{X}_\mathcal{T} \cap \{\boldsymbol{\mu} \in \mathbb{R}^d : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\}$. We now show that $\boldsymbol{v}$ can be represented as a convex combination of points $\boldsymbol{\mu} \in \mathcal{X}_\mathcal{T}$ where $\boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}$ for each $\boldsymbol{\mu}$ in the combination. On the one hand, since $\boldsymbol{v} \in \text{conv } \mathcal{X}_\mathcal{T}$ we can write it as a convex combination $\boldsymbol{v} = \sum_{\boldsymbol{\mu}} \alpha_{\boldsymbol{\mu}} \boldsymbol{\mu}$ for $\boldsymbol{\mu} \in \mathcal{X}_\mathcal{T}$, $\alpha_{\boldsymbol{\mu}} \in [0, 1]$ where we assume without loss of generality that the sum contains only $\alpha_{\boldsymbol{\mu}} > 0$. On the other hand, since $\boldsymbol{v} \in \{\boldsymbol{\mu} \in \mathbb{R}^d : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\}$, it implies that $\sum_{\boldsymbol{\mu}} \alpha_{\boldsymbol{\mu}} \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}$ must hold. To prove the subset relationship it suffices to show that $\mu_i = \mu^*_i$ for all $i \in \mathcal{I}$ for each of the points $\boldsymbol{\mu}$ in the convex combination. Now let $\mu^*_i = 1$ for some $i \in \mathcal{I}$. Assuming that there is a $\hat{\boldsymbol{\mu}}$ in our combination with $\hat{\mu}_i = 0$ results in the following contradiction:

$$\sum_{\boldsymbol{\mu}} \alpha_{\boldsymbol{\mu}} \mu_i = \alpha_{\hat{\boldsymbol{\mu}}} \cdot 0 + \sum_{\boldsymbol{\mu} \neq \hat{\boldsymbol{\mu}}} \alpha_{\boldsymbol{\mu}} \underbrace{\mu_i}_{\leqslant 1} \leqslant \sum_{\boldsymbol{\mu} \neq \hat{\boldsymbol{\mu}}} \alpha_{\boldsymbol{\mu}} < 1 = \mu^*_i,$$

that is, $\sum_{\boldsymbol{\mu}} \alpha_{\boldsymbol{\mu}} \mu_i \neq \mu^*_i$. Analogously, considering the case $\mu^*_i = 0$ and assuming the existence of one $\hat{\mu}_i = 1$ gives rise to the following contradiction:

$$\sum_{\boldsymbol{\mu}} \alpha_{\boldsymbol{\mu}} \mu_i = \alpha_{\hat{\boldsymbol{\mu}}} \cdot 1 + \underbrace{\sum_{\boldsymbol{\mu} \neq \hat{\boldsymbol{\mu}}} \alpha_{\boldsymbol{\mu}} \mu_i}_{\geqslant 0} \geqslant \alpha_{\hat{\boldsymbol{\mu}}} > 0 = \mu^*_i,$$

that is, $\sum_{\boldsymbol{\mu}} \alpha_{\boldsymbol{\mu}} \mu_i \neq \mu^*_i$.

" $\supseteq$ " : This direction follows directly from $\{\boldsymbol{\mu} \in \mathcal{X}_\mathcal{T} : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\} \subseteq \mathcal{X}_\mathcal{T}$ and $\{\boldsymbol{\mu} \in \mathcal{X}_\mathcal{T} : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\} \subseteq \{\boldsymbol{\mu} \in \mathbb{R}^d : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\}$ where $\{\boldsymbol{\mu} \in \mathbb{R}^d : \boldsymbol{\mu}_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}\}$ is a convex set.

□

A Proof of Theorem 1

The following derivations imply the existence of a set of assignments $\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^m$ for the individual subproblems according to the statement in the theorem:

$$\begin{aligned} \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1}, \dots, \boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m}} \mathcal{L}(\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^m, \bar{\boldsymbol{u}}) &\overset{L3}{=} \inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1}, \dots, \boldsymbol{\mu}^m \in L_{\mathcal{T}_m}} \mathcal{L}(\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^m, \bar{\boldsymbol{u}}) \\ &= \inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1}, \dots, \boldsymbol{\mu}^m \in L_{\mathcal{T}_m} :\, \boldsymbol{\mu}^j_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}} \mathcal{L}(\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^m, \bar{\boldsymbol{u}}) \\ &\overset{L4}{=} \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1}, \dots, \boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m} :\, \boldsymbol{\mu}^j_\mathcal{I} = \boldsymbol{\mu}^*_\mathcal{I}} \mathcal{L}(\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^m, \bar{\boldsymbol{u}}) \end{aligned}$$

The first equality holds due to Lemma 3. The second equality is due to the following fact. Since strong duality holds for OP in (8), every optimal primal solution is a minimiser of the Lagrangian $\mathcal{L}(\cdot, ..., \cdot, \bar{\boldsymbol{u}})$. Therefore, the set of feasible solutions restricted by the constraints $\boldsymbol{\mu}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}$ contains at least one optimal solution $\boldsymbol{\mu}^1 := \boldsymbol{\mu}^*|_{\mathcal{T}_1}, ..., \boldsymbol{\mu}^m := \boldsymbol{\mu}^*|_{\mathcal{T}_m}$, where $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ denotes a projection to a subspace corresponding to a tree $\mathcal{T}_j$. The third equality can be shown using Lemma 4 as follows:

$$
\begin{aligned}
&\inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1}, ..., \boldsymbol{\mu}^m \in L_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \mathcal{L}(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_m, \bar{\boldsymbol{u}}) \\
&= \inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1}, ..., \boldsymbol{\mu}^m \in L_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \left\{ \sum_{j=1}^m (\boldsymbol{\theta}_j + \bar{\boldsymbol{u}}_j)^\top \boldsymbol{\mu}^j \right\} \\
&= \inf_{\boldsymbol{\mu}^1 \in L_{\mathcal{T}_1} \,:\, \boldsymbol{\mu}^1_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \{(\boldsymbol{\theta}_1 + \bar{\boldsymbol{u}}_1)^\top \boldsymbol{\mu}^1\} + ... + \inf_{\boldsymbol{\mu}^m \in L_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^m_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \{(\boldsymbol{\theta}_m + \bar{\boldsymbol{u}}_m)^\top \boldsymbol{\mu}^m\} \\
&\overset{(a)}{=} \inf_{\boldsymbol{\mu}^1 \in \mathcal{M}_{\mathcal{T}_1} \,:\, \boldsymbol{\mu}^1_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \{(\boldsymbol{\theta}_1 + \bar{\boldsymbol{u}}_1)^\top \boldsymbol{\mu}^1\} + ... + \inf_{\boldsymbol{\mu}^m \in \mathcal{M}_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^m_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \{(\boldsymbol{\theta}_m + \bar{\boldsymbol{u}}_m)^\top \boldsymbol{\mu}^m\} \\
&\overset{(b)}{=} \inf_{\boldsymbol{\mu}^1 \in \text{ conv } \{\boldsymbol{\mu} \in \mathcal{X}_{\mathcal{T}_1} \,:\, \boldsymbol{\mu}_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}\}} \{(\boldsymbol{\theta}_1 + \bar{\boldsymbol{u}}_1)^\top \boldsymbol{\mu}^1\} + ... + \inf_{\boldsymbol{\mu}^m \in \text{ conv } \{\boldsymbol{\mu} \in \mathcal{X}_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}\}} \{(\boldsymbol{\theta}_m + \bar{\boldsymbol{u}}_m)^\top \boldsymbol{\mu}^m\} \\
&\overset{(c)}{=} \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1} \,:\, \boldsymbol{\mu}^1_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \{(\boldsymbol{\theta}_1 + \bar{\boldsymbol{u}}_1)^\top \boldsymbol{\mu}^1\} + ... + \inf_{\boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^m_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \{(\boldsymbol{\theta}_m + \bar{\boldsymbol{u}}_m)^\top \boldsymbol{\mu}^m\} \\
&= \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1}, ..., \boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \left\{ \sum_{j=1}^m (\boldsymbol{\theta}_j + \bar{\boldsymbol{u}}_j)^\top \boldsymbol{\mu}^j \right\} \\
&= \inf_{\boldsymbol{\mu}^1 \in \mathcal{X}_{\mathcal{T}_1}, ..., \boldsymbol{\mu}^m \in \mathcal{X}_{\mathcal{T}_m} \,:\, \boldsymbol{\mu}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}} \mathcal{L}(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_m, \bar{\boldsymbol{u}})
\end{aligned}
$$

where the step in (a) holds because for every tree-structured MRF $\mathcal{T}_j$ the marginal polytope $\mathcal{M}_{\mathcal{T}_j}$ coincides with the local consistency polytope $L_{\mathcal{T}_j}$; in step (b) we use $\mathcal{M}_{\mathcal{T}_j} = \text{conv } \mathcal{X}_{\mathcal{T}_j}$ and Lemma 4; finally, the step in (c) holds because a linear objective over a polytope always achieves its optimum at least at one of the extreme points (that is, corners) of the latter.

□

Note that the agreement on the integral part holds also for edge marginals, that is, for every dimensions in $\boldsymbol{\mu}^*$ with an integral value, even if the nodes of an edge are fractional. Namely, we ca extend the constraint $\bar{\boldsymbol{\mu}}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}$ in Theorem 1 to every dimension having an integral value and the proof still works.

B Proof of Lemma 2

We denote by $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ a projection of a solution $\boldsymbol{\mu}^*$ over a graph $\mathcal{G}$ to a subspace corresponding to a subtree $\mathcal{T}_j$. Since strong duality holds for the problem (8) any primal optimal is a minimiser of the Lagrangian. Therefore, each restriction $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ is a minimiser of a corresponding subproblem over tree $\mathcal{T}_j$. Furthermore, Theorem 1 guarantees an existence of a minimiser $\bar{\boldsymbol{\mu}}^j$ that agrees with the integral part of $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ and differs from $\boldsymbol{\mu}^*$ only on the fractional entries. Note that this holds also for edge marginals. Namely, in the proof of Theorem 1 we can extend the constraints $\bar{\boldsymbol{\mu}}^j_{\mathcal{I}} = \boldsymbol{\mu}^*_{\mathcal{I}}$ to every dimension in $\boldsymbol{\mu}^*$ having an integral value (including edge marginals) and the proof still works. Any point on the line through these two solutions ($\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ and $\bar{\boldsymbol{\mu}}^j$) is also optimal since the corresponding objective is linear. We now show an existence of a corresponding solution $\hat{\boldsymbol{\mu}}^j$ by construction. We define

$$
\hat{\mu}^j_i(x_i) := \begin{cases} \mu^*_i(x_i), & \text{if } i \in \mathcal{I} \\ 1 - \bar{\mu}^j_i(x_i), & \text{otherwise.} \end{cases} \tag{B.1}
$$

for each $i \in \mathcal{V}_j$ and

$$
\hat{\mu}^j_{i,k}(x_i, x_k) := \begin{cases} \mu^*_{i,k}(x_i, x_k), & \text{if } \mu^*_{i,k}(x_i, x_k) \in \{0, 1\} \\ 1 - \bar{\mu}^j_{i,k}(x_i, x_k), & \text{if } \mu^*_{i,k}(x_i, x_k) = 0.5 \end{cases} \tag{B.2}
$$

for each $(i, k) \in \mathcal{E}_j$. It is easy to see that the above definition of $\hat{\boldsymbol{\mu}}^j$ satisfies the equation (12). Furthermore, $\hat{\boldsymbol{\mu}}^j$ lies on the line through $\bar{\boldsymbol{\mu}}^j$ and the restriction $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ and is therefore optimal. We now show that it is feasible, that is, $\hat{\boldsymbol{\mu}}^j \in \mathcal{X}_{\mathcal{T}_j}$. Note that the variables $\hat{\boldsymbol{\mu}}^j$ are either equal to the variables in $\boldsymbol{\mu}^*_{\mathcal{I}}$ or have an opposite value to the variables in $\bar{\boldsymbol{\mu}}^j$. Therefore, they inherit the integrality constraints as well as the normalisation constraints from $\boldsymbol{\mu}^*$ and $\bar{\boldsymbol{\mu}}^j$. Similar argument can be used for the marginalisation constraints. More precisely, for the integral edges, where both end nodes are integral, the marginalisation constraints hold true. We only need to check the cases with non integral edges. This can be done by considering all the cases listed in Lemma 5. We show exemplary one case – the remaining cases are straightforward. In the following we drop the superscript j denoting the subproblem and write only $\hat{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\mu}}$. Consider the case (a) from Lemma 5. First, we have $\mu^*_{i,j}(0, 0) = \mu^*_{i,j}(1, 1) = 0.5$ and $\mu^*_{i,j}(0, 1) = \mu^*_{i,j}(1, 0) = 0$. That is, $\bar{\mu}_{i,j}(0, 1) = \bar{\mu}_{i,j}(1, 0) = 0$. Without loss of generality assume $\bar{\mu}_{i,j}(0, 0) = 1$ and $\bar{\mu}_{i,j}(1, 1) = 0$, that is, $\bar{\mu}_i(0) = \bar{\mu}_j(0) = 1$ and $\bar{\mu}_i(1) = \bar{\mu}_j(1) = 0$. Due to the construction in (B.1) we get $\hat{\mu}_i(0) = \hat{\mu}_j(0) = 0$ and $\hat{\mu}_i(1) = \hat{\mu}_j(1) = 1$. This corresponds to $\hat{\mu}_{i,j}(0,0) = \hat{\mu}_{i,j}(0,1) = \hat{\mu}_{i,j}(1,0) = 0$ and $\hat{\mu}_{i,j}(1,1) = 1$ which is exactly what we get from construction in (B.2). Therefore, we get a valid labelling of an edge and the marginalisation constraints are satisfied. Finally, note that the equality $\boldsymbol{\mu}^*|_{\mathcal{T}_j} = \frac{1}{2}(\bar{\boldsymbol{\mu}}^j + \hat{\boldsymbol{\mu}}^j)$ also holds (including the edge marginals).

□

C Proof of Theorem 2

Let $\boldsymbol{\mu}^*$ be a unique optimal solution of the LP relaxation (5). We use the notation $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$ to denote a reduction of $\boldsymbol{\mu}^*$ to a corresponding subtree. Theorem 1 guarantees an existence of minimisers $\bar{\boldsymbol{\mu}}^1 \in \mathcal{X}_{\mathcal{T}_1}, ..., \bar{\boldsymbol{\mu}}^m \in \mathcal{X}_{\mathcal{T}_m}$ of a corresponding Lagrangian $\mathcal{L}(\cdot, ..., \cdot, \bar{\boldsymbol{u}})$ where each $\bar{\boldsymbol{\mu}}^j$ agrees with the integral part of $\boldsymbol{\mu}^*|_{\mathcal{T}_j}$. We now assume that there is another minimiser $\hat{\boldsymbol{\mu}}^j$ for the j-th subproblem with $\bar{\boldsymbol{\mu}}_i^j(x_i) \neq \hat{\boldsymbol{\mu}}_i^j(x_i)$ for some $i \in \mathcal{I}$, $x_i \in S$ and show that this assumption leads to a contradiction. We do this by constructing another optimal solution of the LP relaxation different from $\boldsymbol{\mu}^*$.

Assume for simplicity a decomposition over individual edges. We consider the following relabelling procedure starting with an edge (i,k) corresponding to the j-th subproblem above. Since $\hat{\boldsymbol{\mu}}^j$ and $\bar{\boldsymbol{\mu}}^j$ both are minimisers for the corresponding subproblem, the average $\tilde{\boldsymbol{\mu}}^j := \frac{1}{2}(\hat{\boldsymbol{\mu}}^j + \bar{\boldsymbol{\mu}}^j)$ is also a minimiser (because the objective is linear) and $\tilde{\boldsymbol{\mu}}_i^j(x_i) = 0.5$ for $x_i \in S$. That is, the i-th node is now assigned with a fractional label 0.5. The remaining nodes x_r $(r \neq k)$ adjacent to x_i can be relabelled in a consistent way to $x_i = 0.5$ such that a corresponding assignment $(0.5, x_r)$ is optimal for the edge (i,r) by using the weak tree agreement property[4]. Namely, since there are two optimal assignments for the edge (i,k) with both values for i, for every adjacent edge (i,k) there must also be optimal assignments with both values for i. Therefore, we can define a new labelling for each edge adjacent to i by computing the average of the corresponding assignments. During this procedure some nodes x_k can change their label. Note that this is possible only for nodes with integral value in $\boldsymbol{\mu}^*$. To validate this claim consider a fractional node x_k. Since x_i is integral in $\boldsymbol{\mu}^*$, there must be (due to lemma 2) optimal assignments $(x_i^*, 0)$ and $(x_i^*, 1)$ for edge (i,k), where x_i^* is the optimal label of x_i according to $\boldsymbol{\mu}^*$. Furthermore, because of $x_i = 0.5$ (due to relabelling $\tilde{\boldsymbol{\mu}}^j$) there also must be an optimal assignment for that edge of the form $(1-x^*, 0)$ or $(1-x^*, 1)$. In any case we can find an optimal average such that $x_i = x_k = 0.5$. That is, the value of fractional x_k does not change!

If a node x_k changes his label to 0.5 during this procedure, we then need to consider all its neighbours (except x_i) and proceed with the relabelling process. More precisely, we have the following cases:
We now consider an edge (i,k) where x_i has been relabelled to 0.5 in previous steps.
Case 1: $I \rightarrow I$
That is, x_i and x_k both have an integral value in $\boldsymbol{\mu}^*$.
(a) x_k does not change by computing a corresponding average, then there is nothing more to do.
(b) x_k changes. We label it with 0.5 and consider all adjacent cases (except x_i).
Case 2: $I \rightarrow F$
That is, x_i is integral in $\boldsymbol{\mu}^*$ and x_k is fractional. There must be (due to lemma 2) optimal assignments $(x_i^*, 0)$ and $(x_i^*, 1)$. Because $x_i = 0.5$ now there must be (due to WTA) an optimal assignment $(1 - x_i^*, 0)$ or $(1 - x_i^*, 1)$ such that a corresponding optimal average results in $x_i = x_k = 0.5$. So the label of x_k does not change.
Case 3: $I \rightarrow I/F$
That is, x_k has an integral value in $\boldsymbol{\mu}^*$ but has been relabelled to 0.5 previously. Due to the WTA there are always assignments such that a corresponding average results in $x_i = x_k = 0.5$.

Since only integral nodes can change their label during the above relabelling procedure, there are no other cases to consider. The relabelling procedure terminates with a new consistent jont labelling $\tilde{\boldsymbol{\mu}}$ different from $\boldsymbol{\mu}^*$. We can prove the statement for arbitrary tree decompositions (not only over edges) by using similar arguments.

□

D On the fractional solutions of LP relaxation

For binary pairwise MRFs the LP relaxation has the property that in every (extreme) optimal solution each fractional node is half integral [3, 17]. Furthermore, each edge marginal is either integral or has fractional values. More precisely, an edge marginal is integral only if both end nodes are integral. In fact, there are six further cases for fractional edge marginals as specified in the following lemma.

Lemma 5 *Let $\boldsymbol{\mu} \in L_G$ be an extreme point. Then each edge marginal $\mu_{i,j}(x_i, x_j)$ is either integral (if both end nodes x_i and x_j are integral) or*
(a) is equal to

[4] Optimal assignments obtained via DD are known to satisfy the weak tree agreement (WTA) condition [12]. In particular, for our purposes we use the following fact. Consider any two trees $\mathcal{T}_i$ and $\mathcal{T}_j$ which share a node x_k. Then for any optimal configuration $\boldsymbol{\mu}^i$ there exists an optimal configuration $\boldsymbol{\mu}^j$ with $\mu_k^i(x_k) = \mu_k^j(x_k)$.

$\mu_{i,j}(x_i, x_j)$	$x_j = 0$	$x_j = 1$
$x_i = 0$	0.5	0
$x_i = 1$	0	0.5

or

$\mu_{i,j}(x_i, x_j)$	$x_j = 0$	$x_j = 1$
$x_i = 0$	0	0.5
$x_i = 1$	0.5	0

if both x_i and x_j are fractional;
(b) is equal to

$\mu_{i,j}(x_i, x_j)$	$x_j = 0$	$x_j = 1$
$x_i = 0$	0.5	0
$x_i = 1$	0.5	0

or

$\mu_{i,j}(x_i, x_j)$	$x_j = 0$	$x_j = 1$
$x_i = 0$	0	0.5
$x_i = 1$	0	0.5

if x_i is fractional and x_j is integral ($x_j = 0$ on the left and $x_j = 1$ on the right);
(c) is equal to

$\mu_{i,j}(x_i, x_j)$	$x_j = 0$	$x_j = 1$
$x_i = 0$	0.5	0.5
$x_i = 1$	0	0

or

$\mu_{i,j}(x_i, x_j)$	$x_j = 0$	$x_j = 1$
$x_i = 0$	0	0
$x_i = 1$	0.5	0.5

if x_j is fractional and x_i is integral ($x_i = 0$ on the left and $x_i = 1$ on the right);

Proof: The integral case is clear. We now assume that a given edge is non integral, that is, at least one of the nodes is fractional. First we show that in every case a matrix corresponding to an edge assignment contains only two different values a and b.
Case (a):
Since every feasible solution $\boldsymbol{\mu} \in L_G$ is subject to the marginalisation constraints $\sum_{x_i} \mu_{i,j}(x_i, x_j) = \mu_j(x_j)$ and $\sum_{x_j} \mu_{i,j}(x_i, x_j) = \mu_i(x_i)$ the following equations must hold

$$\begin{aligned} \mu_{i,j}(0,0) + \mu_{i,j}(0,1) &= \mu_i(0) \\ \mu_{i,j}(1,0) + \mu_{i,j}(1,1) &= \mu_i(1) \\ \mu_{i,j}(0,0) + \mu_{i,j}(1,0) &= \mu_j(0) \\ \mu_{i,j}(0,1) + \mu_{i,j}(1,1) &= \mu_j(1) \end{aligned} \tag{D.1}$$

Due to $\mu_i(0) = \mu_i(1) = \mu_j(0) = \mu_j(1) = 0.5$ it follows from (D.1) that $a := \mu_{i,j}(0,0) = \mu_{i,j}(1,1)$ and $b := \mu_{i,j}(0,1) = \mu_{i,j}(1,0)$. Now we argue that $a, b \in \{0, 0.5\}$. For this purpose assume that the edge marginal $\mu_{i,j}(x_i, x_j)$ contains other than half-integral values. So w.l.o.g. let $a \in (0, 0.5)$, then also $b \in (0, 0.5)$ (otherwise $a + b \neq 0.5$). We now define two different feasible solutions $\boldsymbol{\mu}^1$ and $\boldsymbol{\mu}^2$ which have the same entries as $\boldsymbol{\mu}$ except the entries for the marginal $\mu_{i,j}(x_i, x_j)$, which we define for $\boldsymbol{\mu}^1$ by $a_1 := a + \epsilon$, $b_1 := b - \epsilon$ and for $\boldsymbol{\mu}^2$ by $a_2 := a - \epsilon$ and $b_2 := b + \epsilon$, where ϵ is small enough such that $a_1, a_2, b_1, b_2 \in (0, 0.5)$. Furthermore, due to $a_1 + b_1 = a_2 + b_2 = 0.5$ a corresponding edge assignment is feasible, and therefore the solutions $\boldsymbol{\mu}^1$, $\boldsymbol{\mu}^2$. Since $\boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}^1 + \boldsymbol{\mu}^2)$, the solution $\boldsymbol{\mu}$ is a convex combination of two different feasible solutions, and is therefore not extreme contradicting our assumption that $\boldsymbol{\mu}$ is a corner of the local polytope. So it must hold $a, b \in \{0, 0.5\}$. Finally, $a = 0$ implies $b = 0.5$ and vice versa due to $a + b = 0.5$. The remaining cases in (b) and (c) can be dealt with by using similar arguments as above.

□