Non-Procedural Facade Parsing: Bidirectional Alignment via Linear Programming Supplementary Material

Mateusz Koziński, Guillaume Obozinski and Renaud Marlet

Université Paris-Est, LIGM (UMR CNRS 8049), ENPC
F-77455 Marne-la-Vallée

Contents of this document

This supplementary material contains a derivation of the update equations used in the algorithm and a proof that the slaves resulting from the proposed decomposition have feasible sets with only integral vertices.

1 The formulation

An exhaustive description of the formulation and the associated notation can be found in the paper. We briefly summarize it here for the convenience of the reader.

Table 1. Symbols used in supplementary material

Sets and notation referring to the tree	I	set of indices of image rows
	J	set of indices of image columns
	$\mathcal{I}$	set if indices of image pixels
	K	set of row-classes assigned to rectangles resulting from horizontal splits and split vertically; they are assigned to image rows
	L	set of column-classes assigned to rectangles resulting from vertical splits and split horizontally; they are assigned to image columns
	r	a class corresponding to entire image, and represented by a root of the tree of classes,
	$\mathring{K}$	set of row-classes that have children and a parent $\mathring{K} = K \setminus (T \cup r)$
	$\mathring{L}$	set of col-classes that have children and a parent $\mathring{L} = L \setminus (T \cup r)$
	$\tilde{K}$	set of nonterminal row-classes $\tilde{K} = K \setminus T$
	$\tilde{L}$	set of nonterminal col-classes $\tilde{L} = L \setminus T$
	T	set of terminal rectangle classes; they are assigned to image pixels; note that $\forall t \in T$ it holds $t \in K$ or $t \in L$
	SK	the set of pairs of classes $(k, k') \in K^2$ that are siblings in the tree
	SL	the set of pairs of classes $(l, l') \in L^2$ that are siblings in the tree
	$Pa(n)$	parent of class n
	$Ch(n)$	set of children classes of class n
	$Sib(n)$	set of classes that share a parent with class n
	$Anc(n)$	set of classes that are ancestors of class n in the tree, including n itself
	$Desc(n)$	set of classes that are descendants of n in the tree, including n itself
Costs	c_{ijt}	cost of assigning a pixel class t to pixel (i, j)
	$c_{ikk'}$	pairwise cost for assigning row-classes k and k' to rows i and $i + 1$
	$c_{jll'}$	pairwise cost for assigning col-classes l and l' to columns j and $j + 1$
Variables	z_{ijt}	assignment of terminal class t to pixel (i, j)
	$\mathbf{z}_{ij}$	the vector (z_{ijt}) for a given pixel (i, j) and all $t \in T$
	$\mathbf{z}$	the vector of (z_{ijt}) for all $(i, j) \in \mathcal{I}$ and all $t \in T$
	y_{ik}	assignment of row-class k to image row i
	$\mathbf{y}_i$	the vector of (y_{ik}) for a given row i
	$\mathbf{y}$	the vector of (y_{ik}) for all $i \in I$ and all $k \in K$
	x_{jl}	assignment of col-class l to image column j
	$\mathbf{x}_j$	the vector of (x_{jl}) for a given column j
	$\mathbf{x}$	the vector of all (x_{jl}) for all $j \in J$ and all $j \in L$
	$y_{ikk'}$	assignment of row-class k to row i and class k' to row $i + 1$
	$\tilde{\mathbf{y}}$	the vector of $y_{ikk'}$ for all $i \in K$ and all $(k, k') \in SK$
	$x_{jll'}$	assignment of col-class l to column j and class l' to column $j + 1$
	$\tilde{\mathbf{x}}$	the vector of $x_{jll'}$ for all $j \in L$ and all $(l, l') \in SL$

The objective defined in the paper is

$$E(\mathbf{z},\mathbf{y},\mathbf{x}) = \sum_{(i,j)\in\mathcal{I}} \sum_{t\in T} z_{ijt} c_{ijt} + \sum_{i=1}^{h-1} \sum_{k,k'\in SK} y_{ikk'} c_{kk'} + \sum_{j=1}^{w-1} \sum_{l,l'\in SL} x_{jll'} c_{ll'} \ . \tag{1}$$

Finding the most likely parse is formulated as

$$\min_{\mathbf{z},\mathbf{y},\mathbf{x}} E(\mathbf{z},\mathbf{y},\mathbf{x}) \ , \tag{2}$$

subject to constraints on binary domain of the variables

$$z_{ijt}, y_{ik}, x_{jl} \in \{0,1\} \ , \tag{3}$$

the requirement of single class per pixel

$$\forall (i,j) \in \mathcal{I}, \quad \sum_{t\in T} z_{ijt} = 1 \ , \tag{4}$$

the constraints determining pixel class given classes assigned to row and column

$$\forall (i,j) \in \mathcal{I}, \forall k \in K \qquad \sum_{t\in Desc(k)} z_{ijt} \leq y_{ik} \ , \tag{5a}$$

$$\forall (i,j) \in \mathcal{I}, \forall l \in L \qquad \sum_{t\in Desc(l)} z_{ijt} \leq x_{jl} \ , \tag{5b}$$

the constraints on hierarchical structure of labels assigned to rows and columns

$$\forall i \in I, \forall l \in \mathring{L}, \qquad \sum_{k'\in Ch(l)} y_{ik'} = y_{iPa(l)} \ , \tag{6a}$$

$$\forall j \in J, \forall k \in \mathring{K}, \qquad \sum_{l'\in Ch(k)} x_{jl'} = x_{jPa(k)} \ , \tag{6b}$$

$$\forall i \in I \qquad \sum_{k\in Ch(r)} y_{ik} = 1 \ , \tag{6c}$$

$$\forall j \in J \qquad x_{jr} = 1 \ , \tag{6d}$$

and the constraints defining pairs of classes assigned to neighboring rows and columns

$$\forall i \in \{1,\dots h-1\}, \forall k \in K \sum_{k'\in Sib(k)} y_{ikk'} = y_{ik} \ , \tag{7a}$$

$$\forall i \in \{1,\dots h-1\}, \forall k' \in K \sum_{k'\in Sib(k)} y_{ikk'} = y_{i+1k'} \ , \tag{7b}$$

$$\forall j \in \{1,\dots w-1\}, \forall l \in L \sum_{l'\in Sib(l)} x_{jll'} = x_{jl} \ , \tag{7c}$$

$$\forall j \in \{1,\dots w-1\}, \forall l' \in L \sum_{l'\in Sib(l)} x_{jll'} = x_{j+1l'} \ . \tag{7d}$$

2 Decomposition of the objective

As stated in the paper we propose to relax the constraint (3) on binary domain of the variables and apply the Dual Decomposition algorithm to the resulting problem. In this section we present a decomposition of the objective (1) into a sum of slave objectives.

As detailed in the paper, in order to represent minimization of the objective (1) subject to constraints (4) to (7) as a sum of tractable slaves we transform constraints (6) to

$$\forall l \in \tilde{L} \quad \sum_{k \in H_l} y_{ik} = 1 \ , \tag{8a}$$

$$\forall k \in \tilde{K} \quad \sum_{l \in V_k} x_{jl} = 1 \ , \tag{8b}$$

where $\tilde{L} = L \setminus T$ and $\tilde{K} = K \setminus T$ are the sets of nonterminal column- and row-classes, $V_k = Ch(k) \cup [L \cap (Ch(A_k) \setminus A_k)]$ and $H_l = Ch(l) \cup [K \cap (Ch(A_l) \setminus A_l)]$, $A_l = Anc(l)$ and $Ch(A_l)$ denotes the set of all children of all elements of A_l.

We decompose the objective (1) into one component for each nonterminal column-class $l \in \tilde{L}$ and one for each nonterminal row-class $k \in \tilde{K}$. Following the scheme of Dual Decomposition, we introduce a separate set of variables for each of the subproblems, and we denote the variables with superscripts l and k, respectively. We note that y^l_{ik} is only defined for $k \in H_l$, and $y^l_{ikk'}$ is defined only for $(k, k') \in SH_l$, where SH_l the set of pairs of sibling row classes k, k' such that $k, k' \in H_l$. Similarly, we introduce variables x^k_{jl} for each $l \in V_k$ and $x^k_{jll'}$ for $(l, l') \in SV_k$, where SV_k is the set of sibling column-classes that belong to V_k. We denote the number of times the pair (k, k') appears in sets SH_l for all $l \in \tilde{L}$ by $n_{kk'}$ and the number of times (l, l') appears in SV_k for different $k \in \tilde{K}$ by $n_{ll'}$. We denote the vector created by stacking y_{ik} by $\mathbf{y}$, the vector of y^l_{ik} by $\mathbf{y}^l$, and the vectors obtained by stacking x_{jl} and x^k_{jl} by $\mathbf{x}$ and $\mathbf{x}^k$. We denote the vectors of z_{ijt}, z^l_{ijt} and z^k_{ijt} by $\mathbf{z}$, $\mathbf{z}^l$ and $\mathbf{z}^k$, respectively. We introduce the components of the new objective

$$E^l(\mathbf{z}^l, \mathbf{y}^l) = \sum_{\substack{(i,j) \in \mathcal{I} \\ t \in T}} \frac{c_{ijt}}{|\tilde{K}| + |\tilde{L}|} z^l_{ijt} + \sum_{i=1}^{h-1} \sum_{(k,k') \in SH_l} y^l_{ikk'} \frac{c_{kk'}}{n_{kk'}} \ , \tag{9a}$$

$$E^k(\mathbf{z}^k, \mathbf{x}^k) = \sum_{\substack{(i,j) \in \mathcal{I} \\ t \in T}} \frac{c_{ijt}}{|\tilde{K}| + |\tilde{L}|} z^k_{ijt} + \sum_{j=1}^{w-1} \sum_{(l,l') \in SV_k} x^k_{jll'} \frac{c_{ll'}}{n_{ll'}} \ . \tag{9b}$$

We abuse the notation by omitting the vectors of pairwise variables $y^l_{ikk'}$ and $x^k_{jll'}$ in the list of arguments of E^l and E^k, because in our setting they will be completely determined by $\mathbf{y}^l$ and $\mathbf{x}^k$. The new objective becomes

$$E(\mathbf{z}^k, \mathbf{x}^k, \mathbf{z}^l, \mathbf{y}^l) = \sum_{k \in \tilde{K}} E^k(\mathbf{z}^k, \mathbf{x}^k) + \sum_{l \in \tilde{L}} E^l(\mathbf{z}^l, \mathbf{y}^l) \ . \tag{10}$$

Our optimization problem now consists in minimizing the new objective

$$\min_{\mathbf{z}^k,\mathbf{x}^k,\mathbf{z}^l,\mathbf{y}^l} E(\mathbf{z}^k,\mathbf{x}^k,\mathbf{z}^l,\mathbf{y}^l) \tag{11}$$

subject to the non-negativity constraints on the relaxed variables:

$$\forall l \in \tilde{L} \quad \mathbf{z}^l \geq 0, \quad \mathbf{y}^l \geq 0 \ , \tag{12a}$$

$$\forall k \in \tilde{K} \quad \mathbf{z}^k \geq 0, \quad \mathbf{x}^k \geq 0 \ , \tag{12b}$$

the coupling constraints

$$\forall k \in \tilde{K} \quad \mathbf{z}^k = \mathbf{z} \ , \tag{13a}$$

$$\forall l \in \tilde{L} \quad \mathbf{z}^l = \mathbf{z} \ , \tag{13b}$$

$$\forall i \in I \ \ \forall l \in \tilde{L} \ \ \forall k \in H_l \quad y^l_{ik} = y_{ik} \ , \tag{13c}$$

$$\forall j \in J \ \forall k \in \tilde{K} \ \ \forall l \in V_k \quad x^k_{jl} = x_{jl} \ , \tag{13d}$$

the constraints (8), defining the structure of the segmentation, on the new variables

$$\forall i \in I, \ \forall l \in \tilde{L} \quad \sum_{k \in H_l} y^l_{ik} = 1 \ , \tag{14a}$$

$$\forall j \in J, \ \forall k \in \tilde{K} \quad \sum_{l \in V_k} x^k_{jl} = 1 \ , \tag{14b}$$

the constraints (4) and (5) on the newly introduced variables

$$\forall (i,j) \in \mathcal{I}, \ \forall l \in \tilde{L} \quad \sum_{t \in T} z^l_{ijt} = 1 \ , \tag{15a}$$

$$\forall (i,j) \in \mathcal{I}, \ \forall k \in \tilde{K} \quad \sum_{t \in T} z^k_{ijt} = 1 \ , \tag{15b}$$

$$\forall (i,j) \in \mathcal{I}, \ \forall l \in \tilde{L}, \ \forall k \in H_l \quad \sum_{t \in Desc(k)} z^l_{ijt} \leq y^l_{ik} \ , \tag{16a}$$

$$\forall (i,j) \in \mathcal{I}, \ \forall k \in \tilde{K}, \ \forall l \in V_k \quad \sum_{t \in Desc(l)} z^k_{ijt} \leq x^k_{jl} \ , \tag{16b}$$

and the constraints on the pairwise variables

$$\forall i \in \{1, \dots h-1\}, \ \forall l \in \tilde{L}, \ \forall k \in H_l \sum_{k' \in Sib^l(k)} y^l_{ikk'} = y^l_{ik} \ , \tag{17a}$$

$$\forall i \in \{1, \dots h-1\}, \ \forall l \in \tilde{L}, \ \forall k' \in H_l \sum_{k' \in Sib^l(k)} y^l_{ikk'} = y^l_{i+1k'} \ , \tag{17b}$$

$$\forall j \in \{1, \dots w-1\}, \ \forall k \in \tilde{K}, \ \forall l \in V_k \sum_{l' \in Sib^k(l)} x^k_{jll'} = x^k_{jl} \ , \tag{17c}$$

$$\forall j \in \{1, \dots w-1\}, \ \forall k \in \tilde{K}, \ \forall l' \in V_k \sum_{l' \in Sib^k(l)} x^k_{jll'} = x^k_{j+1l'} \ , \tag{17d}$$

where $Sib^l(k) = \{k'|(k,k') \in SK^l\}$ denotes the set of sibling of class k that belong to the set K^l.

We note that the domains of objective components $E^l(\mathbf{z}^l, \mathbf{y}^l)$ and $E^k(\mathbf{z}^k, \mathbf{x}^k)$ are entangled only by the coupling constraints (13).

2.1 Formulation of the dual problem

In this chapter we present a dual formulation of problem defined by equations (11) to (17).

We denote the vector created by stacking all $(\mathbf{z}^l, \mathbf{z}^k, \mathbf{z})$ by $\mathbf{z}^\star$. Similarly by $\mathbf{y}^\star$ we denote all $y^l_{ik}, y^l_{ikk'}, y_{ik}, y_{ikk'}$ and the vector of all $x^k_{jl}, x^k_{jll'}, x_{jl}, x_{jll'}$ is denoted by $\mathbf{x}^\star$. We introduce the dual variables λ^l_{ijt} and λ^k_{ijt}, corresponding to the constraint (13), and denote the vector resulting from stacking them together by $\boldsymbol{\lambda}$, and the vectors obtained by stacking variables with the same superscripts by $\boldsymbol{\lambda}^l$ and $\boldsymbol{\lambda}^k$. We also introduce dual variables γ^l_{ik} and γ^k_{jl}, corresponding to the constraint (13), and denote their vectors by $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^l$ and $\boldsymbol{\gamma}^k$.

The application of the dual decomposition method to objective (11) consists in formulating its Lagrangian with respect to the coupling constraints (13):

$$\begin{aligned} L_D(\boldsymbol{\lambda}, \boldsymbol{\gamma}) = \min_{\mathbf{z}^\star, \mathbf{y}^\star, \mathbf{x}^\star} & \sum_{k \in \tilde{K}} E^k(\mathbf{z}^k, \mathbf{x}^k) + \sum_{l \in \tilde{L}} E^l(\mathbf{z}^l, \mathbf{y}^l) \\ & + \sum_{l \in \tilde{L}} \langle \boldsymbol{\lambda}^l, (\mathbf{z}^l - \mathbf{z}) \rangle + \sum_{k \in \tilde{K}} \langle \boldsymbol{\lambda}^k, (\mathbf{z}^k - \mathbf{z}) \rangle \\ & + \sum_{l \in \tilde{L}} \langle \boldsymbol{\gamma}^l, (\mathbf{y}^l - \mathbf{y}) \rangle + \sum_{k \in \tilde{K}} \langle \boldsymbol{\gamma}^k, (\mathbf{x}^k - \mathbf{x}) \rangle \ , \end{aligned} \tag{18}$$

subject to constraints (12) and (14) to (17), where by $\langle \cdot, \cdot \rangle$ we denote the inner product. The variables $\mathbf{z}$, $\mathbf{y}$ and $\mathbf{x}$ can be eliminated by ensuring that

$$\sum_{l \in \tilde{L}} \boldsymbol{\lambda}^l + \sum_{k \in \tilde{K}} \boldsymbol{\lambda}^k = 0 \ , \quad \sum_{l \in \tilde{L} \text{ s.t. } k \in H_l} \boldsymbol{\gamma}^l = 0 \quad \text{and} \quad \sum_{k \in \tilde{K} \text{ s.t. } l \in V_k} \boldsymbol{\gamma}^k = 0 \quad , \tag{19}$$

since if (19) does not hold, the minimum in $\mathbf{z}$, $\mathbf{y}$ and $\mathbf{x}$ is infinite. By eliminating $\mathbf{z}$, $\mathbf{y}$ and $\mathbf{x}$ from (18) we can create the modified Lagrangian which decomposes into independent minimizations for each $k \in \tilde{K}$ and $l \in \tilde{L}$:

$$\tilde{L}_D(\boldsymbol{\lambda}, \boldsymbol{\gamma}) = \sum_{k \in \tilde{K}} \tilde{L}^k_D(\boldsymbol{\lambda}^k, \boldsymbol{\gamma}^k) + \sum_{l \in \tilde{L}} \tilde{L}^l_D(\boldsymbol{\lambda}^l, \boldsymbol{\gamma}^l) \ , \tag{20}$$

where

$$\tilde{L}^k_D(\boldsymbol{\lambda}^k, \boldsymbol{\gamma}^k) = \min_{\mathbf{z}^k, \mathbf{x}^k, \tilde{\mathbf{x}}^k} E^k(\mathbf{z}^k, \mathbf{x}^k) + \langle \boldsymbol{\lambda}^k, \mathbf{z}^k \rangle + \langle \boldsymbol{\gamma}^k, \mathbf{x}^k \rangle \ , \tag{21}$$

$$\tilde{L}^l_D(\boldsymbol{\lambda}^l, \boldsymbol{\gamma}^l) = \min_{\mathbf{z}^l, \mathbf{y}^l, \tilde{\mathbf{y}}^l} E^l(\mathbf{z}^l, \mathbf{y}^l) + \langle \boldsymbol{\lambda}^l, \mathbf{z}^l \rangle + \langle \boldsymbol{\gamma}^l, \mathbf{y}^l \rangle \ , \tag{22}$$

and both minimizations are subject to constraints (12) and (14) to (17).

The final form of the dual problem is

$$\max_{\boldsymbol{\lambda},\boldsymbol{\gamma}} \sum_{k\in\tilde{K}} \tilde{L}_D^k(\boldsymbol{\lambda}^k,\boldsymbol{\gamma}^k) + \sum_{l\in\tilde{L}} \tilde{L}_D^l(\boldsymbol{\lambda}^l,\boldsymbol{\gamma}^l)\ , \tag{23}$$

subject to constraint (19). We solve the problem by means of a projected subgradient ascent procedure presented in the paper. In each iteration we update the dual variables $\boldsymbol{\lambda}^l$, $\boldsymbol{\lambda}^k$, $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^l$ and $\boldsymbol{\gamma}^k$ by making a step in the direction of the subgradient and reprojecting them into the feasible set, where the constraint (19) is satisfied. In the next chapter we derive the update equations.

2.2 Dual Decomposition applied to the problem

We solve (23) by a projected subgradient method. It can be shown that the subgradient of (23) with respect to the dual variables contains the following vectors

$$\nabla_{\boldsymbol{\lambda}^k} L_D(\boldsymbol{\lambda},\boldsymbol{\gamma}) \ni \hat{\mathbf{z}}^k,\ \nabla_{\boldsymbol{\lambda}^l} L_D(\boldsymbol{\lambda},\boldsymbol{\gamma}) \ni \hat{\mathbf{z}}^l,\ \nabla_{\boldsymbol{\gamma}^k} L_D(\boldsymbol{\lambda},\boldsymbol{\gamma}) \ni \hat{\mathbf{x}}^k,\ \nabla_{\boldsymbol{\gamma}^l} L_D(\boldsymbol{\lambda},\boldsymbol{\gamma}) \ni \hat{\mathbf{y}}^l \tag{24a}$$

where

$$(\hat{\mathbf{z}}^k,\hat{\mathbf{x}}^k)=\arg\min_{\mathbf{z}^k,\mathbf{x}^k} E^k(\mathbf{z}^k,\mathbf{x}^k)+\langle\boldsymbol{\lambda}^k,\mathbf{z}^k\rangle+\langle\boldsymbol{\gamma}^k,\mathbf{x}^k\rangle \quad \text{and} \tag{25}$$

$$(\hat{\mathbf{z}}^l,\hat{\mathbf{y}}^l)=\arg\min_{\mathbf{z}^l,\mathbf{y}^l} E^l(\mathbf{z}^l,\mathbf{y}^l)+\langle\boldsymbol{\lambda}^l,\mathbf{z}^l\rangle+\langle\boldsymbol{\gamma}^l,\mathbf{y}^l\rangle\ , \tag{26}$$

where the minimizations are subject to constraints (12) and (14) to (17).

Dual Decomposition adapted to the dual problem (23) takes the form presented in algorithm 1. We denote the values of variables at iteration n with a superscript. The dual variables are updated by a step along the subgradient direction and reprojection to the feasible set, where constraints (19) are satisfied. The reprojection consists in subtracting the average from the corresponding variables. We denote the average of all slave solutions at iteration n by

$$\bar{z}_{ijt}^n = \frac{1}{|\tilde{L}|+|\tilde{K}|}\left(\sum_{l\in\tilde{L}} \hat{z}_{ijt}^{l,n} + \sum_{k\in\tilde{K}} \hat{z}_{ijt}^{k,n}\right)\ . \tag{27}$$

$$\bar{y}_{ik}^n = \frac{1}{|L(k)|}\sum_{l'\in L(k)} \hat{y}_{ik}^{l',n}\ , \tag{28}$$

$$\bar{x}_{jl}^n = \frac{1}{|K(l)|}\sum_{k'\in K(l)} \hat{x}_{jl}^{k',n}, \tag{29}$$

where by $L(k)$ we denote a set of $l\in\tilde{L}$ such that $k\in H_l$, and by $K(l)$ a set of $k\in\tilde{K}$ such that $l\in V_k$.

Algorithm 1 Dual Decomposition applied to the problem

$\forall l \in \tilde{L},\ \ \boldsymbol{\lambda}_l^0 \leftarrow 0,\ \boldsymbol{\gamma}_l^0 \leftarrow 0$
$\forall k \in \tilde{K},\ \boldsymbol{\lambda}_k^0 \leftarrow 0,\ \boldsymbol{\gamma}_k^0 \leftarrow 0$
$n \leftarrow 1$
while not converged **do**
$\quad \forall k \in \tilde{K} \quad (\hat{\mathbf{z}}^{k,n+1}, \hat{\mathbf{x}}^{k,n+1}) \leftarrow \arg\min_{\mathbf{z}^k,\mathbf{x}^k} E^k(\mathbf{z}^k, \mathbf{x}^k) + \langle \boldsymbol{\lambda}^{k,n}, \mathbf{z}^k \rangle + \langle \boldsymbol{\gamma}^{k,n}, \mathbf{x}^k \rangle$
$\quad \forall l \in \tilde{L} \quad (\hat{\mathbf{z}}^{l,n+1}, \hat{\mathbf{y}}^{l,n+1}) \leftarrow \arg\min_{\mathbf{z}^l,\mathbf{y}^l} E^l(\mathbf{z}^l, \mathbf{y}^l) + \langle \boldsymbol{\lambda}^{l,n}, \mathbf{z}^l \rangle + \langle \boldsymbol{\gamma}^{l,n}, \mathbf{y}^l \rangle$
$\quad \forall k \in \tilde{K},\ \forall t \in T, \forall (i,j) \in \mathcal{I} \quad \lambda_{ijt}^{k,n+1} \leftarrow \lambda_{ijt}^{k,n} + \alpha_{n+1}\left(\hat{z}_{ijt}^{k,n+1} - \bar{z}_{ijt}^{n+1}\right)$
$\quad \forall l \in \tilde{L},\ \forall t \in T, \forall (i,j) \in \mathcal{I} \quad \lambda_{ijt}^{l,n+1} \leftarrow \lambda_{ijt}^{l,n} + \alpha_{n+1}\left(\hat{z}_{ijt}^{l,n+1} - \bar{z}_{ijt}^{n+1}\right)$
$\quad \forall k \in \tilde{K},\ \forall l \in V_k, \forall j \in J \quad \gamma_{jl}^{k,n+1} \leftarrow \gamma_{jl}^{k,n} + \alpha_{n+1}\left(\hat{x}_{jl}^{k,n+1} - \bar{x}_{jl}^{n+1}\right)$
$\quad \forall l \in \tilde{L},\ \forall k \in H_l, \forall i \in I \quad \gamma_{ik}^{l,n+1} \leftarrow \gamma_{ik}^{l,n} + \alpha_{n+1}\left(\hat{y}_{ik}^{l,n+1} - \bar{y}_{ik}^{n+1}\right)$
$\quad n \leftarrow n+1$
end while
$\hat{\mathbf{z}}, \hat{\mathbf{y}}, \hat{\mathbf{x}} \leftarrow \textsc{Heuristics}(\mathbf{z}^\star, \mathbf{y}^\star, \mathbf{x}^\star, \boldsymbol{\lambda}, \boldsymbol{\gamma})$

3 The slave subproblem

Below we present the structure of a slave subproblem (26) for some l. The slaves for k are created symmetrically. The copies of the variables are denoted with superscripts l. New cost coefficients $\tilde{c}_{ijt} = \frac{c_{ijt}}{(|\tilde{L}|+|\tilde{K}|)}$ and $\tilde{c}_{kk'} = \frac{c_{kk'}}{n_{kk'}}$ are introduced, where $n_{kk'}$ is the number of times the pair k, k' appears in different slaves. The resulting objective is

$$\min_{\mathbf{z}^l,\mathbf{y}^l} E^l(\mathbf{z}^l, \mathbf{y}^l) = \min_{\mathbf{z}^l,\mathbf{y}^l} \sum_{\substack{(i,j)\in\mathcal{I} \\ t\in T}} (\tilde{c}_{ijt} + \lambda_{ijt}^l) z_{ijt}^l + \sum_{\substack{i\in I \\ k\in H_l}} \gamma_{ik}^l y_{ik}^l + \sum_{\substack{i\in\{1,\dots h-1\} \\ (k,k')\in SH_l}} \tilde{c}_{kk'} y_{ikk'}^l \quad , \tag{30}$$

where λ_{ijt}^l is a Lagrange multiplier corresponding to a constraint coupling the variables z_{ijt}^l for different slaves and γ_{ik}^l is a Lagrange multiplier coupling the variables y_{ik}^l for different slaves. The feasible set of the slave problem is defined by constraints (12) and (14) to (17) which have variables with superscript l in their domain. We rewrite the constraints here for future reference:

$$\forall (i,j) \in \mathcal{I},\ \forall t \in T,\ z_{ijt}^l \geq 0\ , \qquad \forall (i,j) \in \mathcal{I}, \quad \sum_{t\in T} z_{ijt}^l = 1\ , \tag{31a}$$

$$\forall i \in I,\ \forall k \in H_l,\ y_{ik}^l \geq 0\ , \qquad \forall i \in I, \quad \sum_{k\in H_l} y_{ik}^l = 1\ , \tag{31b}$$

$$\forall (i,j) \in \mathcal{I}, \quad \forall k \in H_l, \qquad \sum_{t \in Desc(k)} z^l_{ijt} \leq y^l_{ik} \ , \tag{31c}$$

$$\forall i \in \{1, \dots h-1\}, \quad \forall k \in H_l, \qquad \sum_{k' \in Sib^l(k)} y^l_{ikk'} = y^l_{ik} \ , \tag{31d}$$

$$\forall i \in \{1, \dots h-1\}, \quad \forall k' \in H_l, \qquad \sum_{k \in Sib^l(k')} y^l_{ikk'} = y^l_{i+1k'} \ . \tag{31e}$$

3.1 Integrality of the Slave Subproblem

In this section we show that all vertices of the feasible set defined by constraints (31) are integral. We notice that constraints (31b), (31d) and (31e) form a feasible set of relaxation of a binary linear program for finding the most probable configuration of a Markov Chain. Since relaxations of tree-structured graphical models are tight, it is enough to show that the objective remains linear in y^l_{ik} after marginalizing out z^l_{ijt}.

To eliminate z^l_{ijt} from the slave objective (30) we rewrite the objective as

$$\min_{z^l_{ijt}, y^l_{ik}, y^l_{ikk'}} \sum_{i \in I} g(\mathbf{y}^l_i) + \sum_{\substack{i \in I \\ k \in H_l}} \gamma^l_{ik} y^l_{ik} + \sum_{\substack{i \in \{1, \dots h-1\} \\ k,k' \in SH_l}} c^l_{kk'} y^l_{ikk'} \quad , \tag{32}$$

subject to constraints (31b), (31d) and (31e). We recall that by $\mathbf{y}^l_i$ we denote a vector $(y^l_{ik})_{k \in H_l}$ for given i. We argue that $g()$ is a linear function. It can be formulated as

$$g(\mathbf{y}^l_i) = \min_{z^l_{ijt}} \sum_{\substack{j \in J \\ t \in T}} (\tilde{c}_{ijt} + \lambda^l_{ijt}) z^l_{ijt} \tag{33}$$

subject to

$$\forall j \in J, \ \forall t \in T, \ z^l_{ijt} \geq 0 \ , \tag{34}$$

$$\forall j \in J, \quad \sum_{t \in T} z^l_{ijt} = 1 \ , \tag{35}$$

$$\forall j \in J, \quad \forall k \in H_l, \qquad \sum_{t \in Desc(k)} z^l_{ijt} \leq y^l_{ik} \ . \tag{36}$$

We define the complement of the sets of terminal symbols that are descendants of classes $k \in H_l$ as $T^c_l = T \setminus \bigcup_{k \in H_l} Desc(k)$. We denote $\hat{c}^l_{ijt} = \tilde{c}_{ijt} + \lambda^l_{ijt}$ and define the following costs

$$\forall k \in H_l, \quad \mathring{c}_{ijk} = \min_{t \in Desc(k)} \hat{c}^l_{ijt}, \quad t^k_{ij} = \arg\min\nolimits_{t \in Desc(k)} \hat{c}^l_{ijt}, \tag{37}$$

$$\mathring{c}^c_{ij} = \min_{t \in T^c_l} \hat{c}^l_{ijt}, \quad t^c_{ij} = \arg\min\nolimits_{t \in T^c_l} \hat{c}^l_{ijt}. \tag{38}$$

For future reference let us also note that from (31a) and (31b) we have

$$\sum_{k \in H_l} y^l_{ik} = 1 = \sum_{t \in T} z^l_{ijt} \ , \tag{39}$$

which gives

$$\sum_{t \in T_l^c} z_{ijt}^l = \sum_{k \in H_l} \delta_k \ , \quad \text{where} \quad \delta_k = y_{ik}^l - \sum_{t \in Desc(k)} z_{ijt}^l \geq 0 \ , \tag{40}$$

Where the last inequality follows directly from (31c).

Lemma 1. *The optimal argument of the problem (33) to (36) is*

$$z_{ijt}^l = 0 \qquad \textit{if} \quad (t \in Desc(k)) \wedge ((t \neq t_{ij}^k) \vee (\mathring{c}_{ijk} > \mathring{c}_{ij}^c)) \tag{41}$$

$$z_{ijt}^l = y_{ik}^l \qquad \textit{if} \quad (t = t_{ij}^k) \wedge (\mathring{c}_{ijk} \leq \mathring{c}_{ij}^c) \tag{42}$$

$$z_{ijt}^l = 0 \qquad \textit{if} \quad t \in T_l^c \setminus \{t_{ij}^c\} \tag{43}$$

$$z_{ijt}^l = \sum_{k \ s.t. \ \mathring{c}_{ijk} > \mathring{c}_{ij}^c} y_{ik}^l \qquad \textit{if} \quad t = t_{ij}^c \tag{44}$$

Proof. We prove the lemma by contradiction.

To prove (41) we assume the optimal $z_{ijt}^l = \epsilon > 0$ for some $t \in Desc(k) \setminus \{t_{ij}^k\}$ and $z_{ijt_{ij}^k}^l = \epsilon' \geq 0$. By (37) an argument that yields lower value of the objective is $z_{ijt}^l = 0$ and $z_{ijt_{ij}^k}^l = \epsilon' + \epsilon$. It is easy to verify that the new solution is feasible. For the case when $t = t_{ij}^k$ and $\mathring{c}_{ijk} > \mathring{c}_{ij}^c$ assume the optimal $z_{ijt}^l = \epsilon > 0$ and $z_{ijt_{ij}^c}^l = \epsilon' \geq 0$. By the assumption of (41) a better argument can be constructed by setting $z_{ijt}^l = 0$ and $z_{ijt_{ij}^c}^l = \epsilon' + \epsilon$.

To prove (42) assume the optimal $z_{ijt_{ij}^k}^l = y_{ik}^l - \epsilon$, for $\epsilon \geq 0$. We have by (40) and by (41) $\epsilon = \delta_k$. Since according to (40) δ_k is distributed over z_{ijt}^l for $t \in T_l^c$ the cost of positive epsilon is $c_\epsilon \geq (\mathring{c}_{ij}^c - \mathring{c}_{ijk})$. By the assumption of (42) the cost associated to any $t \in T_l^c$ is larger than $\mathring{c}_{ijk}$, so $c_\epsilon > 0$ and the optimal value of $\epsilon = 0$.

We prove that any argument that violates (43) is suboptimal by constructing a better value of the argument in which the value previously assigned to z_{ijt}^l for any $t \in T_l^c \setminus \{t_{ij}^c\}$ is assigned to $z_{ijt_{ij}^c}^l$, which, according to (38) has a lower cost associated to it.

To prove (44) suppose the optimal value of $z_{ijt_{ij}^c}^l = \sum_{k \text{ s.t. } \mathring{c}_{ijk} > \mathring{c}_{ij}^c} y_{ik}^l - \epsilon$ where $\epsilon > 0$. If ϵ is distributed over z_{ijt}^l for $t \in T_l^c \setminus \{t_{ij}^c\}$ then moving it to $z_{ijt_{ij}^c}^l$ will give better energy by (38). If ϵ is distrubuted over z_{ijt}^l such that $t \in Desc(k)$ for k s.t. $\mathring{c}_{ijk} > \mathring{c}_{ij}^c$ then by the definition of the last set moving the ϵ to $z_{ijt_{ij}^c}^l$ will give lower energy value. $\square$

From lemma 1 it follows directly that

$$g(\mathbf{y}_i^l) = \sum_{k \in H_l} y_{ik}^l \left(\sum_{j \in J} \min\{\mathring{c}_{ijk}, \mathring{c}_{ij}^c\} \right) . \tag{45}$$

This function is linear in y_{ik}^l, which concludes the proof that the slave problem has integral vertices.