Supplementary material for the paper “Efficient Training for Pairwise or Higher Order CRFs via Dual Decomposition”

Nikos Komodakis
University of Crete
http://www.csd.uoc.gr/~komod

Abstract

The supplementary material contains technical proofs for theorems in the main paper.

1. Technical proofs

Theorem 2. *Loss $\mathcal{F}_{\{G_i^k\}}$ upper bounds loss $\mathcal{F}_0$,* i.e.*, $\mathcal{F}_0 \leq \mathcal{F}_{\{G_i^k\}}$*

Proof. By definition (16) it holds that

$$\mathcal{F}_0 = \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} L_{G^k}(\bar{\mathbf{x}}^k, \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k; \mathbf{w}) \tag{39}$$

$$\stackrel{(15)}{=} \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \big(\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) - \mathrm{MRF}_{G^k}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)\big) \tag{40}$$

$$\leq \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \Big(\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) - \mathrm{DUAL}_{\{G_i^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)\Big) \tag{41}$$

$$= \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \left(\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) - \max_{\{\boldsymbol{\theta}^{(i,k)}\}} \sum_i \mathrm{MRF}_{G_i^k}(\boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k)\right) \tag{42}$$

$$= \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \min_{\{\boldsymbol{\theta}^{(i,k)}\}} \left(\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) - \sum_i \mathrm{MRF}_{G_i^k}(\boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k)\right) \tag{43}$$

$$= \min_{\mathbf{w}, \{\boldsymbol{\theta}^{(i,k)}\}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \sum_i \Big(\mathrm{MRF}_{G_i^k}(\bar{\mathbf{x}}^k; \boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k) - \mathrm{MRF}_{G_i^k}(\boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k)\Big) \tag{44}$$

$$= \min_{\mathbf{w}, \{\boldsymbol{\theta}^{(i,k)}\}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \sum_i L_{G_i^k}(\bar{\mathbf{x}}^k, \boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k; \mathbf{w}) \ , \tag{45}$$

where inequality (41) is true because $\mathrm{DUAL}_{\{G_i^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ is a convex relaxation of minimization problem $\mathrm{MRF}_{G^k}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ and so it holds $\mathrm{DUAL}_{\{G_i^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) \leq \mathrm{MRF}_{G^k}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$, while equality (44) is satisfied due to the fact that it holds $\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) = \sum_i \mathrm{MRF}_{G_i^k}(\bar{\mathbf{x}}^k; \boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k)$ since $\sum_{i \in \mathcal{I}_p^k} \theta_p^{(i,k)}(\cdot) = \bar{u}_p^k(\cdot)$. □

Theorem 3. *If $\{G_i^k\} < \{\tilde{G}_j^k\}$ then $\mathcal{F}_{\{\tilde{G}_j^k\}} < \mathcal{F}_{\{G_i^k\}}$.*

Proof. By definition (22) it holds that

$$\mathcal{F}_{\{G_i^k\}} = \min_{\mathbf{w},\{\boldsymbol{\theta}^{(i,k)}\}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \sum_i L_{G_i^k}(\bar{\mathbf{x}}^k, \boldsymbol{\theta}^{(i,k)}, \bar{\mathbf{h}}^k; \mathbf{w}) \tag{46}$$

$$= \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \Big(\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) - \mathrm{DUAL}_{\{G_i^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) \Big) \quad , \tag{47}$$

where the equality (47) is derived using a similar reasoning as in the proof of theorem 2 above.

Similarly, the following equality can be shown to hold true

$$\mathcal{F}_{\{\tilde{G}_j^k\}} = \min_{\mathbf{w},\{\boldsymbol{\theta}^{(j,k)}\}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \sum_j L_{\tilde{G}_j^k}(\bar{\mathbf{x}}^k, \boldsymbol{\theta}^{(j,k)}, \bar{\mathbf{h}}^k; \mathbf{w}) \tag{48}$$

$$= \min_{\mathbf{w}} \mu R(\mathbf{w}) + \sum_{k=1}^{K} \Big(\mathrm{MRF}_{G^k}(\bar{\mathbf{x}}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) - \mathrm{DUAL}_{\{\tilde{G}_j^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) \Big) \quad . \tag{49}$$

By assumption it also holds $\{G_i^k\} < \{\tilde{G}_j^k\}$, which means that the convex relaxation $\mathrm{DUAL}_{\{\tilde{G}_j^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ is tighter than the convex relaxation $\mathrm{DUAL}_{\{G_i^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$, which in turn implies that

$$\mathrm{DUAL}_{\{G_i^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) < \mathrm{DUAL}_{\{\tilde{G}_j^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k) \quad . \tag{50}$$

The theorem now follows directly by combining equations (47), (49) and (50). □

Theorem 4. *$\mathcal{F}_{\{\tilde{G}_j^k\}}$ can be a better approximation to $\mathcal{F}_0$ than $\mathcal{F}_{G^k_{\text{single}}}$ only if there exists at least one sub-hypergraph $\tilde{G}_j^k$ such that slave MRFs on $\tilde{G}_j^k$ do not have the integrality property*[1].

Proof. The MRF optimization problem $\mathrm{MRF}_{G^k}(\mathbf{x}^k; \bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ is equivalent to the following linear integer program:

$$\min_{\mathbf{z}} \sum_{p \in \mathcal{V}^k} \sum_{x_p} \bar{u}_p^k(x_p) z_p(x_p) + \sum_{c \in \mathcal{C}^k} \sum_{\mathbf{x}_c} \bar{h}_c^k(\mathbf{x}_c) z_c(\mathbf{x}_c) \tag{51}$$

$$\text{s.t } \mathbf{z} \in Z(G^k) \quad . \tag{52}$$

In the above problem, the feasible set $Z(G^k)$ is defined for any hypergraph $G^k = (\mathcal{V}^k, \mathcal{C}^k)$ as

$$Z(G^k) = \left\{ \mathbf{z} \in \bar{Z}(G^k) \mid z_p(\cdot), z_c(\cdot) \in \{0,1\},\ \forall\, p \in \mathcal{V}^k,\ c \in \mathcal{C}^k \right\} \quad , \tag{53}$$

where

$$\bar{Z}(G^k) = \left\{ \mathbf{z} \;\middle|\; \begin{array}{ll} \sum_{x_p} z_p(x_p) = 1, & \forall\, p \in \mathcal{V}^k \\ \sum_{\mathbf{x}_c : x_p = l} z_c(\mathbf{x}_c) = z_p(l), & \forall\, c \in \mathcal{C}^k,\ p \in c \\ z_p(\cdot) \geq 0, z_c(\cdot) \geq 0, & \forall\, p \in \mathcal{V}^k, c \in \mathcal{C}^k \end{array} \right\} \quad .$$

Let $\{\tilde{G}_j^k = (\tilde{\mathcal{V}}_j^k, \tilde{\mathcal{C}}_j^k)\}$ be a hypergraph decomposition of G^k (*i.e.* $\cup \tilde{\mathcal{V}}_j^k = \mathcal{V}^k$, $\cup \tilde{\mathcal{C}}_j^k = \mathcal{C}^k$, $\tilde{\mathcal{C}}_j^k \cap \tilde{\mathcal{C}}_{j'}^k = \emptyset, \forall j \neq j'$) and let $\{\boldsymbol{\theta}^{(j,k)}\}$ be a set of unary potentials for the corresponding slave MRFs chosen such that they satisfy equation (23), *i.e.*

$$\sum_{j \in \mathcal{I}_p^k} \theta_p^{(j,k)}(\cdot) = \bar{u}_p^k(\cdot), \tag{54}$$

[1]We say that an MRF has the integrality property if and only if the corresponding LP relaxation of integer program (34) is tight.

where $\mathcal{I}_p^k = \{j | p \in \tilde{\mathcal{V}}_j^k\}$ (*e.g.* $\boldsymbol{\theta}^{(j,k)}$ can be chosen as $\theta_p^{(j,k)}(\cdot) = \bar{u}_p^k(\cdot)/|\mathcal{I}_p^k|$). Using these potentials, the above linear integer program (51) can be equivalently expressed as

$$\min_{\mathbf{z},\mathbf{z}^j} \sum_j \left(\sum_{p \in \tilde{\mathcal{V}}_j^k} \sum_{x_p} \theta_p^{(j,k)}(x_p) z_p^j(x_p) + \sum_{c \in \tilde{\mathcal{C}}_j^k} \sum_{\mathbf{x}_c} \bar{h}_c^k(\mathbf{x}_c) z_c^j(\mathbf{x}_c) \right) \tag{55}$$

$$\text{s.t. } \mathbf{z}^j \in Z(\tilde{G}_j^k) \ , \ \forall j \tag{56}$$

$$z_p^j(\cdot) = z_p(\cdot) \, , \ \forall p \in \mathcal{V}^k \ . \tag{57}$$

The convex relaxation $\mathrm{DUAL}_{\{\tilde{G}_j^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ is derived by relaxing constraints (57) and then solving the resulting Lagrangean relaxation. Therefore, $\mathrm{DUAL}_{\{\tilde{G}_j^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ is equivalent to the following relaxation of the above integer progam

$$\min_{\mathbf{z},\mathbf{z}^j} \sum_j \left(\sum_{p \in \tilde{\mathcal{V}}_j^k} \sum_{x_p} \theta_p^{(j,k)}(x_p) z_p^j(x_p) + \sum_{c \in \tilde{\mathcal{C}}_j^k} \sum_{\mathbf{x}_c} \bar{h}_c^k(\mathbf{x}_c) z_c^j(\mathbf{x}_c) \right) \tag{58}$$

$$\text{s.t. } \mathbf{z}^j \in \mathrm{CH}\left(Z(\tilde{G}_j^k) \right) \ , \ \forall j \tag{59}$$

$$z_p^j(\cdot) = z_p(\cdot) \qquad , \ \forall p \in \mathcal{V}^k \ , \tag{60}$$

where $\mathrm{CH}(A)$ denotes the convex hull of set A. If we now assume that all slave MRFs corresponding to decomposition $\{\tilde{G}_j^k\}$ have the integrality property then by definition this implies that $\mathrm{CH}\left(Z(\tilde{G}_j^k) \right) = \bar{Z}(\tilde{G}_j^k)$ (*i.e.* we can safely ignore the integrality constraints in (53)) and so $\mathrm{DUAL}_{\{\tilde{G}_j^k\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{h}}^k)$ reduces to

$$\min_{\mathbf{z},\mathbf{z}^j} \sum_j \left(\sum_{p \in \tilde{\mathcal{V}}_j^k} \sum_{x_p} \theta_p^{(j,k)}(x_p) z_p^j(x_p) + \sum_{c \in \tilde{\mathcal{C}}_j^k} \sum_{\mathbf{x}_c} \bar{h}_c^k(\mathbf{x}_c) z_c^j(\mathbf{x}_c) \right) \tag{61}$$

$$\text{s.t. } \mathbf{z}^j \in \bar{Z}(\tilde{G}_j^k) \ , \ \forall j \tag{62}$$

$$z_p^j(\cdot) = z_p(\cdot) \, , \ \forall p \in \mathcal{V}^k \ . \tag{63}$$

Due to constraints (54) and (63), the objective function (61) above is easily seen to be equal to the objective function (51), and so the above relaxation is obviously equal to the LP relaxation of (34) corresponding to decomposition G_{single}^k, which concludes the proof of the theorem. $\square$