A note on the primal-dual method for the semi-metric labeling problem

Vladimir Kolmogorov
vnk@adastral.ucl.ac.uk

Technical report
June 14, 2007

Abstract

Recently, Komodakis et al. *[6] developed the* FastPD *algorithm for the semi-metric labeling problem, which extends the expansion move algorithm of Boykov* et al. *[2]. We present a slightly different derivation of the FastPD method.*

1. Preliminaries

Consider the following energy function:

$$E(\mathbf{x} \mid \bar\theta) = \sum_{u\in\mathcal{V}} \bar\theta_u(x_v) + \sum_{(u,v)\in\mathcal{E}} \bar\theta_{uv}(x_u,x_v) \qquad (1)$$

Here $\mathcal{G}=(\mathcal{V},\mathcal{E})$ is an undirected graph. Variables x_u for nodes $u\in\mathcal{V}$ belong to a discrete set $x_u\in\mathcal{X}_u$. Thus, labeling $\mathbf{x}$ belongs to the set $\mathcal{X}=\otimes_{u\in\mathcal{V}}\mathcal{X}_u$.

We denote $\vec{\mathcal{E}}=\{(u\to v),(v\to u)\mid(u,v)\in\mathcal{E}\}$, i.e. $(\mathcal{V},\vec{\mathcal{E}})$ is the directed graph corresponding to undirected graph $(\mathcal{V},\mathcal{E})$.

The energy function (1) is specified by unary terms $\theta_u(i)$ and pairwise terms $\theta_{uv}(i,j)$ ($i\in\mathcal{X}_u, j\in\mathcal{X}_v$). It will be convenient to denote them as $\bar\theta_{u;i}$ and $\bar\theta_{uv;ij}$, respectively. We can concatenate all these values into a single vector $\bar\theta=\{\bar\theta_\alpha\mid\alpha\in\mathcal{I}\}$ where the index set is $\mathcal{I}=\{(u;i)\}\cup\{(uv;ij)\}$. Note that $(uv;ij)\equiv(vu;ji)$, so $\bar\theta_{uv;ij}$ and $\bar\theta_{vu;ji}$ are the same element. We will use the notation $\bar\theta_u$ to denote a vector of size $|\mathcal{X}_u|$ and $\bar\theta_{uv}$ to denote a vector of size $|\mathcal{X}_u\times\mathcal{X}_v|$.

1.1. LP relaxation

In this section we describe a linear programming (LP) relaxation of energy (1) which plays a crucial role for algorithms in [5, 6]. This "natural" relaxation was studied extensively in the literature, in particular by Schlesinger [10] (for a special case when $\theta_{uv}(i,j)\in\{0,+\infty\}$), Koster *et al.* [7], Chekuri *et al.* [3], and Wainwright *et al.* [11].

Primal problem Let us introduce binary indicator variables: $\tau_{u;i}=[x_u=i]$, $\tau_{uv;ij}=[x_u=i,x_v=j]$ where $[\cdot]$ is the Iverson bracket: it is 1 if its argument is true, and 0 otherwise. Variables $\{\tau_{u;i}\},\{\tau_{uv;ij}\}$ must belong to the following constraint set:

$$\Lambda=\left\{\tau\in\mathbb{R}_+^{\mathcal{I}}\;\middle|\;\begin{array}{ll}\sum_{i\in\mathcal{X}_u}\tau_{u;i}=1 & \forall\, u\in\mathcal{V}\\ \sum_{i\in\mathcal{X}_u}\tau_{uv;ij}=\tau_{v;j} & \forall\,(u\to v)\in\vec{\mathcal{E}},\\ & j\in\mathcal{X}_v\end{array}\right\}$$

Clearly, the problem of minimizing function (1) can be formulated as follows:

$$\begin{array}{ll}\text{minimize} & \langle\bar\theta,\tau\rangle\\ \text{subject to} & \tau\in\Lambda\\ & \tau_{uv;ij}\in\{0,1\}\end{array}$$

The LP relaxation in [10, 7, 3, 11, 5, 6] is obtained by dropping the integrality constraint:

$$\begin{array}{ll}\text{minimize} & \langle\bar\theta,\tau\rangle\\ \text{subject to} & \tau\in\Lambda\end{array}\qquad(2a)$$

Reparameterization and dual problem The dual problem to (2a) can be defined using the notion of reparameterization [10, 11].

Definition 1. *Suppose vectors θ and $\bar\theta$ define the same energy function, i.e. $E(\mathbf{x}\mid\theta)=E(\mathbf{x}\mid\bar\theta)$ for all configurations $\mathbf{x}$. Then θ is called a* reparameterization *of $\bar\theta$.*

Let us introduce *message* $M_{uv;j}\in\mathbb{R}$ for directed edge $(u\to v)\in\vec{\mathcal{E}}$ and label $j\in\mathcal{X}_v$. We denote $M_{uv}=\{M_{uv;j}\mid j\in\mathcal{X}_v\}$ to be a vector of size $|\mathcal{X}_v|$, and $M=\{M_{uv}\mid(u\to v)\in\vec{\mathcal{E}}\}$ to be the vector of all messages. This vector defines reparameterization $\theta=\bar\theta[M]$ as follows:

$$\begin{aligned}\theta_{u;i} &= \bar\theta_u+\sum_{(u,v)\in\mathcal{E}} M_{vu;i}\\ \theta_{uv;ij} &= \bar\theta_{uv;ij}-M_{uv;j}-M_{vu;i}\end{aligned}$$

In this paper $\bar{\theta}$ denotes the original parameter vector, and $\theta = \bar{\theta}[M]$ denotes its reparameterization defined by some message vector M. To simplify notation, we denote the corresponding energy function as $E(\mathbf{x}) = E(\mathbf{x}\,|\,\theta) = E(\mathbf{x}\,|\,\bar{\theta})$.

Let us define

$$\Phi(\theta) = \sum_{u\in\mathcal{V}} \min_{i\in\mathcal{X}_u} \theta_{u;i} + \sum_{(u,v)\in\mathcal{E}} \min_{\substack{i\in\mathcal{X}_u\\ j\in\mathcal{X}_v}} \theta_{uv;ij}$$

It is easy to see that for any messages M value $\Phi(\bar{\theta}[M])$ is a lower bound on the energy: $\Phi(\bar{\theta}[M]) \le \min_{\mathbf{x}} E(\mathbf{x})$. This motivates the following maximization problem:

$$\begin{array}{ll}\text{maximize} & \Phi(\theta)\\ \text{subject to} & \theta = \bar{\theta}[M]\end{array} \tag{2b}$$

In other words, the goal is to find the tightest possible bound. This maximization problem is dual to (2a) (see e.g. [12]).

1.2. Optimality conditions

Consider fractional labeling $\tau \in \Lambda$ and messages M corresponding to reparameterization $\theta = \bar{\theta}[M]$. It is well-known that (τ, M) is an optimal primal-dual pair (i.e. τ is an optimal solution of (2a) and M is an optimal solution of (2a)) if and only if the following *complementary slackness* conditions hold:

$$\tau_{u;i} > 0 \quad\Rightarrow\quad \theta_{u;i} = \min_{i'\in\mathcal{X}_u} \theta_{u;i'} \tag{3a}$$

$$\tau_{uv;ij} > 0 \quad\Rightarrow\quad \theta_{uv;ij} = \min_{\substack{i'\in\mathcal{X}_u\\ j'\in\mathcal{X}_v}} \theta_{uv;i'j'} \tag{3b}$$

Algorithms in [5, 6] maintain an **integer** primal vector $\tau \in \Lambda$ (or equivalently a labeling $\mathbf{x} \in \mathcal{X}$). With such a restriction reparameterization $\theta = \bar{\theta}[M]$ satisfying (3) may not exist. Thus, conditions (3) must be relaxed. Given number $f_{app} \ge 1$, let us say that $(\mathbf{x}, M)$ satisfies f_{app}*-relaxed complementary slackness* conditions if the following holds for all nodes and edges:

$$\theta_u(x_u) \le \min_{i\in\mathcal{X}_u} \theta_{u;i} + \left(1 - \frac{1}{f_{app}}\right)\bar{\theta}_u(x_u) \tag{4a}$$

$$\theta_{uv}(x_u, x_v) \le \min_{\substack{i\in\mathcal{X}_u\\ j\in\mathcal{X}_v}} \theta_{uv;ij} + \left(1 - \frac{1}{f_{app}}\right)\bar{\theta}_{uv}(x_u, x_v) \tag{4b}$$

Theorem 2 (cf. [5])**.** *Suppose that pair* $(\mathbf{x}, M)$ *satisfies eq.* (4)*. Then* $\mathbf{x}$ *is an* f_{app}*-approximation, i.e.* $E(\mathbf{x}) \le f_{app} \min_{\mathbf{y}} E(\mathbf{y})$.

Proof. Let us sum (4a) over nodes $u \in \mathcal{V}$ and (4b) over edges $(u,v) \in \mathcal{E}$. We obtain

$$\begin{aligned} E(\mathbf{x}) &\le \Phi(\theta) + \left(1 - \frac{1}{f_{app}}\right) E(\mathbf{x}) \\ \frac{1}{f_{app}} E(\mathbf{x}) &\le \Phi(\theta) \le \min_{\mathbf{y}} E(\mathbf{y}) \end{aligned}$$

□

1.3. LP relaxation for submodular functions of binary variables

The expansion move algorithm [2] and algorithms in [5, 6] rely on solving the minimization problem with (at most) **binary** variables. In this section we consider the case when $\mathcal{X}_u = \{0,1\}$ or $\mathcal{X}_u = \{0\}$ for all nodes u. Furthermore, we assume that function E is submodular, i.e. each term θ_{uv} with $\mathcal{X}_u = \mathcal{X}_v = \{0,1\}$ satisfies $\theta_{uv;00} + \theta_{uv;11} \le \theta_{uv;01} + \theta_{uv;10}$. (Note that expression $\theta_{uv;00} + \theta_{uv;11} - \theta_{uv;01} - \theta_{uv;10}$ is invariant to reparameterization.)

This case has several important properties (see e.g. [1]). First, it can solved efficiently by computing a maximum flow in a graph with $|\mathcal{V}| + 2$ nodes and $|\mathcal{V}| + |\mathcal{E}|$ edges. Second, the algorithm produces an **integer** optimal solution $\tau \in \Lambda$ (or equivalently labeling $\mathbf{x} \in \mathcal{X}$ which is a global minimum of E). Thus, optimality conditions (3) are reduced to

$$\theta_u(x_u) = \min_{i\in\mathcal{X}_u} \theta_{u;i} \tag{5a}$$

$$\theta_{uv}(x_u, x_v) = \min_{\substack{i\in\mathcal{X}_u\\ j\in\mathcal{X}_v}} \theta_{uv;ij} \tag{5b}$$

Finally, reparameterization θ is in a *normal form*, i.e. terms θ_{uv} satisfy

$$\theta_{uv}(0,0) = \theta_{uv}(i_{\max}, j_{\max}) = \min_{\substack{i\in\mathcal{X}_u\\ j\in\mathcal{X}_v}} \theta_{uv;ij} \tag{6}$$

where $i_{\max}$ is the maximal label in $\mathcal{X}_u$ and $j_{\max}$ is the maximal label in $\mathcal{X}_v$.

Handling singleton nodes In a practical implementation nodes u with $\mathcal{X}_u = \{0\}$ can be handled as follows. First, for each “boundary” edge (u,v) with $\mathcal{X}_u = \{0\}, \mathcal{X}_v = \{0,1\}$ we choose message M_{uv} so that $\theta_{uv;00} = \theta_{uv;01}$. Namely, $M_{uv;0} = 0$, $M_{uv;1} = \bar{\theta}_{uv;01} - \bar{\theta}_{uv;00}$. Then we solve the LP relaxation for nodes u with $\mathcal{X}_u = \{0,1\}$ ignoring singleton nodes and their incident edges. In this step only messages M_{uv} for edges $(u \to v)$ with $\mathcal{X}_u = \mathcal{X}_v = \{0,1\}$ are allowed to be modified. Clearly, upon termination conditions (5) and (6) will hold for all nodes and edges.

Restricting messages It is easy to see that adding a constant to vector M_{uv} preserves optimality of M. Thus, when solving problem (2b) we can require that $M_{uv;0} = 0$ for all $(u \to v) \in \vec{\mathcal{E}}$. This constraint will be used later.

Given term $\bar{\theta}_{uv}$ and reparameterization θ_{uv} for edge (u,v) with $\mathcal{X}_u = \mathcal{X}_v = \{0,1\}$, messages M_{uv}, M_{vu} can be computed as follows: (1) add constant $C = \bar{\theta}_{uv;00} - \theta_{uv;00}$ to vector θ_{uv} so that we get $\theta_{uv;00} = \bar{\theta}_{uv;00}$; (2) set $M_{uv;0} = M_{vu;0} = 0$, $M_{uv;1} = \bar{\theta}_{uv;01} - \theta_{uv;01}$, $M_{vu;1} = \bar{\theta}_{uv;10} - \theta_{uv;10}$.

2. FastPD algorithm [6]

From now on we assume that the set $\mathcal{X}_u$ is the same for all nodes: $\mathcal{X}_u = \mathcal{L}$, and that vector $\bar{\theta}$ satisfies the following for all nodes and edges:

$$\bar{\theta}_{u;i} \geq 0 \qquad \forall\, i \in \mathcal{L} \tag{7a}$$
$$\bar{\theta}_{uv;ii} = 0 \qquad \forall\, i \in \mathcal{L} \tag{7b}$$
$$\bar{\theta}_{uv;ij} > 0 \qquad \forall\, i,j \in \mathcal{L}, i \neq j \tag{7c}$$

We also consider a special case when the triangular inequalities hold:

$$\bar{\theta}_{uv;ij} \geq \bar{\theta}_{uv;ik} + \bar{\theta}_{uv;kj} \qquad \forall\, i,j,k \in \mathcal{L} \tag{7d}$$

Note that if we add the symmetry condition ($\bar{\theta}_{uv}(i,j) = \bar{\theta}_{uv}(j,i)$) then eq. (7a)-(7c) give the definition of a semi-metric, and eq. (7a)-(7d) give the definition of a metric, However, the symmetry will not be needed.

We denote $\bar{\theta}^{\min}_{uv} = \min_{i,j\in\mathcal{L},i\neq j} \bar{\theta}_{uv;ij}$, $\bar{\theta}^{\max}_{uv} = \max_{i,j\in\mathcal{L}} \bar{\theta}_{uv;ij}$ and $f_{app} = 2\max_{(u,v)\in\mathcal{E}} \frac{\bar{\theta}^{\max}_{uv}}{\bar{\theta}^{\min}_{uv}}$.

2.1. General overview

FastPD algorithm [6] maintains integer primal configuration $\mathbf{x} \in \mathcal{X}$ and dual variables (messages) M which define reparameterization $\theta = \bar{\theta}[M]$. The following invariants hold during the execution:

$$\theta_{uv}(x_u, x_v) = 0 \tag{8a}$$
$$\theta_{uv}(i,i) = 0 \qquad \forall\, i \in \mathcal{L} \tag{8b}$$

These properties can easily be ensured in the beginning: for any $\mathbf{x}$ it is straightforward to find messages M so that (8) holds. At each step, the algorithm selects some label $k \in \mathcal{L}$ and performs the k-*expansion* operation described in section 2.2. After the first pass over labels in $\mathcal{L}$ the following condition holds for all nodes u:

$$\theta_u(x_u) = \min_{i\in\mathcal{L}} \theta_{u;i} \tag{9}$$

The algorithm terminates when there was a pass over all labels $k \in \mathcal{L}$ but configuration $\mathbf{x}$ has not changed. At this point an additional property holds for all edges. In the case of triangular inequalities (7d) we have

$$\theta_{uv;ij} \geq 0 \qquad \forall\, i,j \in \mathcal{L}, i = x_u \text{ or } j = x_v \tag{10}$$

Without triangular inequalities eq. (10) cannot be guaranteed. Instead, the following weaker version holds:

$$\theta_{uv;ij} \geq \bar{\theta}_{uv;ij} - \bar{\theta}^{\max}_{uv} \qquad \forall\, i,j \in \mathcal{L}, i = x_u \text{ or } j = x_v \tag{10$'$}$$

The final step of the algorithm is to scale messages down:

$$M_{uv;j} := M_{uv;j}/f_{app} \tag{11}$$

After that the f_{app}-relaxed complementary slackness conditions (4) will hold and therefore labeling $\mathbf{x}$ will be within factor f_{app} from the optimum.

2.2. k-expansion step

The input to this step is label k and primal-dual pair $(\mathbf{x}^\circ, M^\circ)$. Let us denote $\widetilde{\mathcal{X}}_u = \{x^\circ_u, k\}$, and let $\widetilde{E}$ be the restriction of function E to configurations $\mathbf{x}$ with $x_u \in \widetilde{\mathcal{X}}_u$. $\widetilde{E}$ can be viewed as an energy function of (at most) binary variables. We assume that label x°_u corresponds to 0 and label k corresponds to 1 (if $k \neq x^\circ_u$). Note that messages M define reparameterization not only for the original energy E, but also for the restriction $\widetilde{E}$.

Submodular case First, let us consider the case when function $\widetilde{E}$ is submodular. (We get this case if, for example, triangular inequalities (7d) hold.) The k-expansion step solves the LP relaxation for the restriction $\widetilde{E}$, as described in section 1.3. The LP relaxation has an integer solution. Thus, the goal is to find a global minimum of $\widetilde{E}$ and messages M that give optimal reparameterization for $\widetilde{E}$:

$$\mathbf{x} := \arg\min_{x_u\in\widetilde{\mathcal{X}}_u} \widetilde{E}(\mathbf{x}) \tag{12a}$$
$$M := \arg\max_{M:\theta=\bar{\theta}[M]} \widetilde{\Phi}(\theta) \tag{12b}$$

where the objective function in the maximization problem is

$$\widetilde{\Phi}(\theta) = \sum_{u\in\mathcal{V}} \min_{i\in\widetilde{\mathcal{X}}_u} \theta_{u;i} + \sum_{(u,v)\in\mathcal{E}} \min_{\substack{i\in\widetilde{\mathcal{X}}_u \\ j\in\widetilde{\mathcal{X}}_v}} \theta_{uv;ij}$$

To achieve efficiency, one could start with reparameterization $\theta^\circ = \bar{\theta}[M^\circ]$ when solving (12b). This can be formulated as follows: (i) find message increment M^δ and corresponding reparameterization $\theta = \theta^\circ[M^\delta]$ which maximizes $\widetilde{\Phi}(\theta)$; (ii) set $M := M^\circ + M^\delta$.

When solving problem (12b), only components $M_{uv;k}$ for edges $(u \to v)$ with $x^\circ_v \neq k$ will be allowed to change. In other words, $M_{uv;j} = M^\circ_{uv;j}$ for labels $j \in \mathcal{L}_v$ where $\mathcal{L}_v = (\mathcal{L} - \{k\}) \cup \{x^\circ_v\}$. This is not a severe restriction: as discussed in section 1.3, it still allows to find an *optimal* solution to (12b).

Function $\widetilde{E}$ may have several global minima. If their cost equals $E(\mathbf{x}^\circ)$ then $\mathbf{x}^\circ$ is not updated, i.e. $\mathbf{x}$ is chosen as $\mathbf{x}^\circ$. This guarantees convergence since the number of configurations is finite.

Note that in the primal domain the method is equivalent the expansion move algorithm of Boykov *et al.* [2].

Non-submodular case If function $\widetilde{E}$ is non-submodular then the following operations are performed. For each non-submodular edge (u,v) with

$$\bar{\theta}_{uv}(x^\circ_u, x^\circ_v) > \bar{\theta}_{uv}(x^\circ_u, k) + \bar{\theta}_{uv}(k, x^\circ_v)$$

terms $\bar{\theta}_{uv}(x^\circ_u, k)$ and $\bar{\theta}_{uv}(k, x^\circ_v)$ are increased by non-negative constants until we get an equality:

$$\bar{\theta}_{uv}(x^\circ_u, x^\circ_v) = \bar{\theta}'_{uv}(x^\circ_u, x^\circ_v) = \bar{\theta}'_{uv}(x^\circ_u, k) + \bar{\theta}'_{uv}(k, x^\circ_v)$$

where $\bar{\theta}'$ is the modified parameter vector. The algorithm then proceeds in the same way as before. Namely, define $\widetilde{E}'$ to be the restriction of function $E(\cdot \mid \bar{\theta}')$ to labelings $\mathbf{x}$ with $x_u \in \widetilde{\mathcal{X}}_u$. It is easy to see that $\widetilde{E}'$ is submodular. The LP relaxation of $\widetilde{E}'$ is now solved yielding pair $(\mathbf{x}, M)$ and corresponding parameter vector $\theta' = \bar{\theta}'[M]$. Finally, vector $\bar{\theta}'$ is restored to its original value $\bar{\theta}$, and vector θ' is changed to $\theta = \bar{\theta}[M]$.

Note that in the primal domain this algorithm is a special case of the more general majorize-minimize technique (see e.g. [8]). It is also equivalent to the "truncation" trick described in [9]. In the context of the labeling problem "truncation" was proposed independently in [4] and in [9]. Both papers state that the method produces an f_{app}-approximation, but in [9] this fact is mentioned without proof.

2.3. Correctness of FastPD

Theorem 3 (cf. [6]). *(i) After the k-expansion reparameterization $\theta = \bar{\theta}[M]$ satisfies conditions* (8)*.*

(ii) After the first pass over all labels k condition (9) *holds.*

(iii) Upon convergence pair $(\mathbf{x}, M)$ satisfies condtion (10) *(in the case of triangular inequalities* (7d)*) or condition* (10′) *(without triangular inequalities).*

Proof. **Submodular case** By assumption, conditions (8) hold for the initial reparameterization $\theta^\circ = \bar{\theta}[M^\circ]$:

$$\theta^\circ_{uv}(x_u, x_v) = 0 \tag{13a}$$
$$\theta^\circ_{uv}(i,i) = 0 \quad \forall\, i \in \mathcal{L} \tag{13b}$$

We have $M_{uv;j} = M^\circ_{uv;j}$ for labels $j \in \mathcal{L}_v$. Therefore,

$$\theta_u(i) = \theta^\circ_u(i) \qquad \forall\, i \in \mathcal{L}_u \tag{13c}$$
$$\theta_{uv}(i,j) = \theta^\circ_{uv}(i,j) \qquad \forall\, i \in \mathcal{L}_u, j \in \mathcal{L}_v \tag{13d}$$

Eq. (13d) and (13a) yield

$$\theta_{uv}(x^\circ_u, x^\circ_v) = 0 \tag{13e}$$

Using eq. (13e) and condition (6) of the normal form we obtain

$$\theta_{uv}(k,k) = 0 \tag{13f}$$
$$\theta_{uv}(x^\circ_u, k) \geq 0, \quad \theta_{uv}(k, x^\circ_v) \geq 0 \tag{13g}$$

Finally, from the optimality conditions (5) we obtain

$$\theta_u(x_u) \leq \theta_u(i) \quad \forall i \in \widetilde{\mathcal{X}}_u \tag{13h}$$
$$\theta_{uv}(x_u, x_u) = 0 \tag{13i}$$

Non-submodular case Let $\theta^\circ = \bar{\theta}[M^\circ]$ and $\theta = \bar{\theta}[M]$ be the reparameterizations before changing $\bar{\theta}$ and after restoring $\bar{\theta}$, respectively. We claim that conditions (13) hold except that (13g) is replaced with the following:

$$\begin{aligned}\theta_{uv}(x^\circ_u, k) &\geq \bar{\theta}_{uv}(x^\circ_u, k) - \bar{\theta}^{\max} \\ \theta_{uv}(k, x^\circ_v) &\geq \bar{\theta}_{uv}(k, x^\circ_v) - \bar{\theta}^{\max}\end{aligned} \tag{13g′}$$

Indeed, eq. (13a)-(13f), (13h)-(13i) can be shown in the same way as before, using the fact that $\theta_{uv}(k,k) = \theta'_{uv}(k,k)$, $\theta'_{uv}(x^\circ_u, x^\circ_v) = \theta_{uv}(x^\circ_u, x^\circ_v)$ where $\theta' = \bar{\theta}'[M]$ is the reparameterization of $\bar{\theta}'$ obtained after solving the LP relaxation. Instead of (13g) we now have

$$\theta'_{uv}(x^\circ_u, k) \geq 0, \quad \theta'_{uv}(k, x^\circ_v) \geq 0$$

If the term for edge (u,v) is submodular then $\theta_{uv} = \theta'_{uv}$, so (13g′) clearly holds. Otherwise,

$$\begin{aligned}M_{vu;x^\circ_u} + M_{uv;k} &= \bar{\theta}'_{uv}(x^\circ_u, k) - \theta'_{uv}(x^\circ_u, k) \\ &\leq \bar{\theta}'_{uv}(x^\circ_u, k) \\ &= \bar{\theta}_{uv}(x^\circ_u, x^\circ_v) - \bar{\theta}'_{uv}(k, x^\circ_v) \\ &\leq \bar{\theta}_{uv}(x^\circ_u, x^\circ_v) \leq \theta^{\max}_{uv}\end{aligned}$$

which implies the first inequality in (13g′). The second inequality follows from the first inequality applied to the reverse edge.

We now prove the theorem using equations (13).

Proof of (i) We already showed (8a) (see (13i)). Eq. (8b) follows from (13b), (13d) and (13f).

Proof of (ii) Let $\widetilde{\mathcal{L}}$ be the set of labels processed so far at least once. Let us prove by induction on the number of steps that $\theta_u(x_u) \leq \theta_u(i)$ for each node $u \in \mathcal{V}$ and label $i \in \widetilde{\mathcal{L}}$.

The base of the induction is trivial: in the beginning $\widetilde{\mathcal{L}}$ is empty. Suppose that the claim holds for vector θ° and labeling $\mathbf{x}^\circ$ in the beginning of the k-expansion step. The fact $\theta_u(x_v) \leq \theta_u(i)$ for label $i = k$ follows from (13h). For label $i \in \widetilde{\mathcal{L}} - \{k\}$ we have

$$\theta_u(x_u) \leq \theta_u(x^\circ_u) = \theta^\circ_u(x^\circ_u) \leq \theta^\circ_u(i) = \theta_u(i)$$

(The first inequality follows from (13h), the second inequality is by the induction hypothesis, and the two equalities follow from (13c)).

Proof of (iii) We consider only the case with triangular inequalities (7d). (The proof for the general case is entirely analogous; we just need to use eq. (13g′) instead of (13g).)

Consider the last iteration, i.e. a complete pass over labels $k \in \mathcal{L}$ in which configuration $\mathbf{x}$ has not changed. Let $\widetilde{\mathcal{L}}$ be the set of labels processed in this iteration. Let us prove by induction on the number of steps that $\theta_{uv}(x,j) \geq 0$ for labels $j \in \widetilde{\mathcal{L}}$.

The base of the induction is trivial: in the beginning $\widetilde{\mathcal{L}}$ is empty. Suppose that the claim holds for vector θ° in the beginning of the k-expansion step. Eq. (13d) and (13g) imply that the claim also holds for vector θ and set $\widetilde{\mathcal{L}} \cup \{k\}$ after the step.

The analogous fact for term $\theta_{uv}(k, x_v)$ can be shown in the same way.

□

Theorem 4 (cf. [5, 6])**.** *If reparameterization* $\theta = \bar{\theta}[M]$ *satisfies conditions* (8), (9) *and* (10′) *then after message scaling* (11) *the* f_{app}*-relaxed slackness conditions* (4) *hold.*

Proof. Let $M' = M/f_{app}$ be the messages after scaling and $\theta' = \bar{\theta}[M']$ be the corresponding reparameterization.

Proof of (4a) Consider node u. There holds

$$\begin{aligned}\theta'_{u;i} &= \bar{\theta}_{u;i} + \sum_{(u,v)\in\mathcal{E}} M'_{vu;i} = \bar{\theta}_{u;i} + \frac{\theta_{u;i} - \bar{\theta}_{u;i}}{f_{app}} \\ &= \frac{1}{f_{app}}\theta_{u;i} + \left(1 - \frac{1}{f_{app}}\right)\bar{\theta}_{u;i}\end{aligned}$$

Using this equality and eq. (9) we obtain

$$\begin{aligned}\theta'_u(x_u) &= \frac{1}{f_{app}} \min_{i\in\mathcal{L}} \theta_{u;i} + \left(1 - \frac{1}{f_{app}}\right)\bar{\theta}_u(x_u) \\ &\leq \min_{i\in\mathcal{L}} \theta'_{u;i} + \left(1 - \frac{1}{f_{app}}\right)\bar{\theta}_u(x_u)\end{aligned}$$

Proof of (4b) Now consider edge (u, v). Let us show that $\theta'_{uv;ij} \geq 0$ for all $i, j \in \mathcal{L}$. If $i = j$ then this follows from (8b). Suppose that $i \neq j$. From (10′) we get

$$\begin{aligned}M_{vu;x_u} + M_{uv;j} &\leq \bar{\theta}^{\max}_{uv} \\ M_{vu;i} + M_{uv;x_v} &\leq \bar{\theta}^{\max}_{uv} \\ M_{vu;x_u} + M_{uv;x_v} &= 0\end{aligned}$$

Adding the first two equations and subtracting the third one yields

$$\begin{aligned}M_{vu;i} + M_{uv;j} &\leq 2\,\bar{\theta}^{\max}_{uv} \\ M'_{vu;i} + M'_{uv;j} = \frac{M_{vu;i} + M_{uv;j}}{f_{app}} &\leq \bar{\theta}^{\min}_{uv}\end{aligned}$$

$$\theta'_{uv;ij} = \bar{\theta}_{uv;ij} - (M'_{vu;i} + M'_{uv;j}) \geq \bar{\theta}_{uv;ij} - \bar{\theta}^{\min}_{uv} \geq 0$$

as claimed. Thus, $\min_{i,j\in\mathcal{L}} \theta'_{uv;ij} \geq 0$.

Finally, using (8a) we get

$$\begin{aligned}\theta'_{uv}(x_u, x_v) &= \bar{\theta}_{uv}(x_u, x_v) - M'_{uv;x_v} - M'_{vu;x_u} \\ &= \bar{\theta}_{uv}(x_u, x_v) - \frac{\bar{\theta}_{uv}(x_u, x_v) - \theta_{uv}(x_u, x_v)}{f_{app}} \\ &= \left(1 - \frac{1}{f_{app}}\right)\bar{\theta}_{uv}(x_u, x_v)\ .\end{aligned}$$

□

Acknowledgements I thank Pushmeet Kohli and Philip Torr for discussions on connections between primal-dual schema and tree-reweighted message passing.

References

[1] E. Boros and P. L. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155 – 225, November 2002.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11), November 2001.

[3] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *SODA*, pages 109–118, 2000.

[4] N. Komodakis and G. Tziritas. Approximate labeling via the primal-dual schema. Technical Report CSD-TR-2005-01, University of Crete, February 2005.

[5] N. Komodakis and G. Tziritas. Approximate labeling via graph-cuts based on linear programming. *PAMI*, 2007. (To appear).

[6] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic MRFs. In *CVPR*, 2007.

[7] A. Koster, C. van Hoesel, and A. Kolen. The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters*, 23(3-5):89–97, 1998.

[8] K. Lange. *Optimization*. Springer Verlag, New York, 2004.

[9] C. Rother, S. Kumar, V. Kolmogorov, and A. Blake. Digital tapestry. In *CVPR*, 2005.

[10] M. I. Schlesinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika*, 4:113–130, 1976.

[11] M.J. Wainwright, T.S. Jaakkola, and A.S. Willsky. MAP estimation via agreement on trees: Message-passing and linear-programming approaches. *IEEE Transactions on Information Theory*, 51(11):3697–3717, November 2005.

[12] T. Werner. A linear programming approach to max-sum problem: A review. *PAMI*, 29(7):1165–1179, July 2007.