What Is Optimized in Tight Convex Relaxations for Multi-Label Problems? Supplementary material

Christopher Zach (MSR Cambridge), Christian Häne (ETH Zürich), Marc Pollefeys (ETH Zürich)

April 2, 2012

1 Notations

We recall the following definitions from the main paper. The original tight relaxation of multi-label problems [Chambolle et al., 2008] reads as

$$E_{\text{CCP-I}}(u,q) = \sum_{s,i} \theta_s^i (u_s^{i+1} - u_s^i) + \sum_{s,i} (q_s^i)^T \nabla u_s^i \tag{1}$$
$$\text{s.t. } u_s^i \le u_s^{i+1},\, u_s^0 = 0,\, u_s^{L+1} = 1,\, u_s^i \ge 0$$
$$\Big\| \sum_{k=i}^{j-1} q_s^k \Big\|_2 \le \theta^{ij} \qquad \forall s,i,j,$$

The corresponding version in terms of node-wise pseudo-marginals is given by

$$E_{\text{CCP-II}}(x,p) = \sum_{s,i} \theta_s^i x_s^i + \sum_{s,i} (p_s^i)^T \nabla x_s^i \tag{2}$$
$$\text{s.t. } \left\| p_s^i - p_s^j \right\|_2 \le \theta^{ij},\, x_s \in \Delta \qquad \forall s,i,j,$$

In the main paper we state the following primal energy of Eq. 2

$$E_{\text{tight}}(x,y) = \sum_{s,i} \theta_s^i x_s^i + \sum_s \sum_{i,j:i<j} \theta^{ij} \|y_s^{ij}\|_2 \tag{3}$$
$$\text{s.t. } \nabla x_s^i = \sum_{j:j<i} y_s^{ji} - \sum_{j:j>i} y_s^{ij}\; x_s \in \Delta \qquad \forall s,i,$$

as well as this one,

$$E(x) = \sum_{s,i} \theta_s^i x_s^i + \sum_s \sum_{i,j:i<j} \theta^{ij} \left\| x_s^{ij} + x_s^{ji} \right\|_2 \tag{4}$$
$$\text{s.t. } \nabla x_s^i = \sum_{j:j\ne i} x_s^{ji} - \sum_{j:j\ne i} x_s^{ij},\, x_s \in \Delta,\, x_s^{ij} \ge 0 \qquad \forall s,i.$$

2 Switching Between Superlevel and Indicator Representations

In this section we show the equivalence between Eq. 1 and Eq. 2. In the main paper we subsequently focus on Eq. 2.

We use u_s^i to denote the superlevel representation and x_s^i for the indicator representation of a label assignment, i.e. $x_s = \partial_i u_s$, where we use backward differences for ∂_i and $u_s^0 = 0$ as boundary condition. With these definitions we obtain $x_s^1 = u_s^1$ and $x_s^i = u_s^i - u_s^{i-1}$, which is desired. We have (in 2 dimensions, but this generalizes to any dimension)

$$\nabla_x \partial_i u_s = \begin{pmatrix} (u^i_{s+(1,0)} - u^{i-1}_{s+(1,0)}) - (u^i_s - u^{i-1}_s) \\ (u^i_{s+(0,1)} - u^{i-1}_{s+(0,1)}) - (u^i_s - u^{i-1}_s) \end{pmatrix} = \begin{pmatrix} (u^i_{s+(1,0)} - u^i_s) - (u^{i-1}_{s+(1,0)} - u^{i-1}_s) \\ (u^i_{s+(0,1)} - u^i_s) - (u^{i-1}_{s+(0,1)} - u^{i-1}_s) \end{pmatrix} = \partial_i \nabla_x u_s.$$

Since we have $x_s = \partial_i u_s$,

$$\max_{p_s \in C} \langle p_s, \nabla_x x_s \rangle = \max_{p_s \in C} \langle p_s, \nabla_x \partial_i u_s \rangle = \max_{p_s \in C} \langle p_s, \partial_i \nabla_x u_s \rangle = \max_{p_s \in C} \langle \partial_i^T p_s, \nabla_x u_s \rangle,$$

where C is the constraint set $C = \{p : \|p^i - p^j\| \le \theta^{ij}\}$. Explicitly we have

$$\partial_i u_s^k = \begin{cases} u_s^1 & \text{if } k = 1 \\ u_s^k - u_s^{k-1} & \text{if } 1 < k \le L. \end{cases}$$

For a 6-label problem the matrix corresponding to ∂_i is

$$\begin{pmatrix} 1 & & & & & \\ -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & -1 & 1 & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{pmatrix}.$$

For the adjoint operator ∂_i^T we have

$$\partial_i^T p_s^k = \begin{cases} p_s^k - p_s^{k+1} & \text{if } 1 \le k < L \\ p_s^L & \text{if } k = L. \end{cases}$$

The solution of $\partial_i^T p_s = q_s$ is of the form $p_s^k = \sum_{l=k}^L q_s^l$ (like an antiderivative), and the constraints expressed in terms of q_s are

$$\theta^{ij} \ge \|p_s^i - p_s^j\| = \Big\| \sum_{l=i}^L q_s^l - \sum_{l=j}^L q_s^l \Big\| = \begin{cases} \big\| \sum_{l=i}^{j-1} q_s^l \big\| & \text{if } i < j \\ \big\| \sum_{l=j}^{i-1} q_s^l \big\| & \text{if } i > j, \end{cases} \tag{5}$$

which are exactly the constraints used in the super-level representation Eq. 1.

3 The Primal of the Isotropic Tight Convex Relaxation

Since $E_{\text{CCP-II}}$ can be written as

$$E_{\text{CCP-II}}(x) = \sum_{s,i} \theta_s^i x_s^i + \sum_s \max_{p_s^i} \sum_i (p_s^i)^T \nabla x_s^i \tag{6}$$
$$\text{s.t. } \big\|p_s^i - p_s^j\big\|_2 \le \theta^{ij},\ x_s \in \Delta,$$

we only need to consider the point-wise problem

$$\max_{p_s^i} \sum_i (p_s^i)^T \nabla x_s^i \qquad \text{subject to } \big\|p_s^i - p_s^j\big\|_2 \le \theta^{ij}. \tag{7}$$

We will omit the subscript s and derive the primal of

$$\max_{p^i} \sum_i (p^i)^T \nabla x^i \qquad \text{subject to } \left\|p^i - p^j\right\|_2 \le \theta^{ij} \ \ \forall i < j.$$

Fenchel duality ($-f^*(-A^T p) - g^*(p) \rightsquigarrow f(y) + g(-Ay)$) leads to the primal

$$\sum_{i,j:i<j} \theta^{ij} \left\|y^{ij}\right\|_2 \qquad \text{subject to } Ay = \nabla x, \tag{8}$$

since the conjugate of $f \equiv \imath\{\|\cdot\|_2 \le \theta\}$ is $\theta\|\cdot\|_2$, and the conjugate of $g \equiv a^T\cdot$ is $\imath\{\cdot = a\}$. The matrix $-A$ (which has rows corresponding to p^i and columns corresponding to y^{ij}) has a -1 entry at position (p^i, y^{ij}) (for $i < j$) and a +1 element at (p^j, y^{ij}) $(i > j)$. Thus, the i-th row of $-Ay$ reads as

$$\sum_{j:j<i} y^{ji} - \sum_{j:j>i} y^{ij}, \tag{9}$$

and the purely primal form of Eq. 7 is given by

$$\begin{aligned} &\min_{y_s^{ij}} \sum_{i,j:i<j} \theta^{ij} \left\|y_s^{ij}\right\|_2 \\ &\text{s.t. } \nabla x_s^i = \sum_{j:j<i} y_s^{ji} - \sum_{j:j>i} y_s^{ij}. \end{aligned} \tag{10}$$

By replacing the inner maximization problem in Eq. 6 with this expression we obtain E_{tight}.

We can express the primal energy also in terms of non-negative pseudo-marginals. We start with the decoupled binary potentials from Eq. 4,

$$\begin{aligned} E_s(x) &= \sum_{i,j:i<j} \theta^{ij} \left\|x_s^{ij} + x_s^{ji}\right\|_2 + \sum_{i,j} \imath\{x_s^{ij} \ge 0\} \\ \text{s.t. } \nabla x_s^i &= \sum_{j:j\ne i} x_s^{ji} - \sum_{j:j\ne i} x_s^{ij}, \end{aligned} \tag{11}$$

and dualize E_s. First, we note that every optimal solution satisfies complementarity constraints $x_s^{ij} \perp x_s^{ji}$, i.e. $(x_s^{ij})_k (x_s^{ji})_k = 0$ (otherwise one could strictly decrease the norm term by setting $x_s^{ij} \leftarrow x_s^{ij} - \min\{x_s^{ij}, x_s^{ij}\}$ and $x_s^{ji} \leftarrow x_s^{ji} - \min\{x_s^{ij}, x_s^{ij}\}$ without affecting the marginalization constraint). Hence, we have

$$\begin{aligned} \left\|x_s^{ij} + x_s^{ji}\right\|_2 &= \sqrt{((x_s^{ij})_1 + (x_s^{ji})_1)^2 + ((x_s^{ij})_2 + (x_s^{ji})_2)^2} \\ &= \sqrt{((x_s^{ij})_1)^2 + ((x_s^{ji})_1)^2 + \underbrace{(x_s^{ij})_1 (x_s^{ji})_1}_{=0} + ((x_s^{ij})_2)^2 + ((x_s^{ji})_2)^2 + \underbrace{(x_s^{ij})_2 (x_s^{ji})_2}_{=0}} \\ &= \left\|((x_s^{ij})_1, (x_s^{ij})_2, (x_s^{ji})_1, (x_s^{ji})_2)^T\right\| = \left\|\begin{matrix} x_s^{ij} \\ x_s^{ji} \end{matrix}\right\|_2. \end{aligned}$$

Consequently, E_s above can be restated as

$$\begin{aligned} E_s(x) &= \sum_{i,j:i<j} \theta^{ij} \left\|\begin{matrix} x_s^{ij} \\ x_s^{ji} \end{matrix}\right\|_2 + \sum_{i,j} \imath\{x_s^{ij} \ge 0\} \\ \text{s.t. } \nabla x_s^i &= \sum_{j:j\ne i} x_s^{ji} - \sum_{j:j\ne i} x_s^{ij}, \end{aligned} \tag{12}$$

which seems to be more convenient to work with. Using the fact that the conjugate of $f(x) = \theta\|x\|_2 + \imath\{x \geq 0\}$ is $f^*(p) = \imath\{[p]_+ \leq \theta\}$ (see Section 4), we obtain the following dual of E_s,

$$E_s^*(p) = \sum_i (p_s^i)^T \nabla x_s^i \qquad \text{s.t.} \ \left\| \begin{matrix} [p_s^i - p_s^j]_+ \\ [p_s^j - p_s^i]_+ \end{matrix} \right\|_2 \leq \theta^{ij} \tag{13}$$

$$= \sum_i (p_s^i)^T \nabla x_s^i \qquad \text{s.t.} \ \left\| p_s^i - p_s^j \right\|_2 \leq \theta^{ij}, \tag{14}$$

which is exactly Eq. 7. Thus, we have shown the equivalence of the primal programs Eq. 3 and Eq. 4.

4 Dual Energies

If we consider the primal energy

$$\begin{aligned} E_{\text{tight}}(x, y) = \sum_{s,i} \theta_s^i x_s^i + \sum_s \sum_{i,j:i<j} \theta^{ij} \|y_s^{ij}\|_2 \qquad \text{subject to} \\ \nabla x_s^i = \sum_{j:j<i} y_s^{ji} - \sum_{j:j>i} y_s^{ij}, \ x_s \in \Delta, \end{aligned} \tag{15}$$

the dual energy is given by

$$E_{\text{tight-I}}^*(p) = \sum_s \min_i \{\operatorname{div} p_s^i + \theta_s^i\} - \sum_s \sum_{i,j:i<j} \imath\{\|p_s^i - p_s^j\|_2 \leq \theta^{ij}\}.$$

Note that we have redundant constraints on the primal variables $y_s^{ij} \in [-1, 1] \times [-1, 1]$ (since $x_s^i \in [0, 1]$). One could compute the dual of $\theta^{ij} \|y_s^{ij}\|_2 + \imath\{\|y_s^{ij}\|_\infty \leq 1\}$, but because of its radial symmetry the constraint $\|y_s^{ij}\|_2 \leq \sqrt{2}$ seems to be more appropriate. Via

$$\left(x \mapsto \theta|x| + \imath_{[0,B]}(x)\right)^*(y) = \max_{x \in [0,B]} \{xy - \theta|x|\} = B \max\{0, |y| - \theta\}$$

and the radial symmetry of terms in y_s^{ij} we obtain for the full dual energy in this setting

$$E_{\text{tight-II}}^*(p) = \sum_s \min_i \{\operatorname{div} p_s^i + \theta_s^i\} + \sum_s \sum_{i,j:i<j} \sqrt{2} \min\left\{0, \theta^{ij} - \|p_s^i - p_s^j\|_2\right\}.$$

The first term in $E_{\text{tight-II}}^*$, $\sum_s \min_i \{\operatorname{div} p_s^i + \theta_s^i\}$, can also be replaced by penalty terms: if we move the normalization constraint $\sum_i x_s^i = 1$ to the linear constraints and introduce a respective Lagrange multiplier q_s, we obtain via

$$\left(x \mapsto \theta x + \imath_{[0,1]}(x)\right)^*(y) = [y - \theta]_+ \qquad \text{and}$$
$$(\imath_{\{x:Ax=b\}})^*(y) = \imath_{\operatorname{im}(A^T)}(y) + b^T\lambda \qquad \text{for } y = A^T\lambda$$

the dual energy in p_s^i and q_s:

$$E_{\text{tight-III}}^*(p, q) = \sum_s q_s + \sum_{s,i} \left[\operatorname{div} p_s^i + \theta_s^i - q_s\right]_- + \sum_s \sum_{i,j:i<j} \sqrt{2} \min\left\{0, \theta^{ij} - \|p_s^i - p_s^j\|_2\right\}, \tag{16}$$

5 Proof of Observation 1

This section shows that for graph-based MRFs with truncated smoothness costs a compact representation is equivalent to the full one. We have the full model,

$$E_{\text{full}} = \sum_{s,i} \theta_s^i x_s^i + \sum_{(s,t)\in\mathcal{E}} \sum_{i,j} \theta^{ij} x_{st}^{ij}$$
$$= \sum_{s,i} \theta_s^i x_s^i + \sum_{(s,t)\in\mathcal{E}} \left(\sum_{i,j:|i-j|<T} \theta^{ij} x_{st}^{ij} + \theta^* \sum_{i,j:|i-j|\geq T} x_{st}^{ij} \right) \tag{17}$$

subject to the marginalization constraints $\sum_j x_{st}^{ij} = x_s^i$ and $\sum_i x_{st}^{ij} = x_t^i$. We assume $\theta^{ij} = \theta^*$ for $|i-j| \geq T$ (where T is the truncation point) and $\theta^{ij} < \theta^*$. The reduced program reads as

$$E_{\text{red}} = \sum_{s,i} \theta_s^i x_s^i + \sum_{(s,t)\in\mathcal{E}} \left(\sum_{i,j:|i-j|<T} \theta^{ij} x_{st}^{ij} + \frac{\theta^*}{2} \sum_i (x_{st}^{i*} + x_{st}^{*i}) \right) \tag{18}$$

with the slightly different marginalization constraints

$$x_s^i = \sum_{i,j:|i-j|<T} x_{st}^{ij} + x_{st}^{i*} \qquad \text{and} \qquad x_t^j = \sum_{i,j:|i-j|<T} x_{st}^{ij} + x_{st}^{*j}.$$

If we have a minimizer of E_{full}, we can easily construct a solution of E_{red} with the same overall objective by setting

$$x_{st}^{i*} = \sum_{j:|i-j|\geq T} x_{st}^{ij} \qquad \text{and} \qquad x_{st}^{*j} = \sum_{i:|i-j|\geq T} x_{st}^{ij},$$

since the pairwise truncated smoothness costs are the same

$$\frac{\theta^*}{2} \sum_i x_{st}^{i*} + \frac{\theta^*}{2} \sum_j x_{st}^{*j} = \frac{\theta^*}{2} \sum_i \sum_{j:|i-j|\geq T} x_{st}^{ij} + \frac{\theta^*}{2} \sum_j \sum_{i:|i-j|\geq T} x_{st}^{ij} = \theta^* \sum_{i,j:|i-j|\geq T} x_{st}^{ij}. \tag{19}$$

If we have a minimizer x of E_{red}, we have to construct a solution $\hat{x}$ of E_{full} with the same objective. We set

$$\hat{x}_s^i = x_s^i \qquad \text{and} \qquad \hat{x}_{st}^{ij} = x_{st}^{ij} \qquad \forall i,j : |i-j| < T.$$

Determining x_{st}^{ij} for $i,j : |i-j| \geq T$ is more difficult. In the following we consider a particular edge st and omit the subscript. We use the north-west corner rule-like to assign $\hat{x}^{ij}$ for $i,j : |i-j| \geq T$:

$\bar{x}^{i*} \leftarrow x^{i*}$
$\bar{x}^{*j} \leftarrow x^{*j}$
while some $\hat{x}^{ij}$ is not assigned **do**
  Choose i and j (with $|i-j| \geq T$) such that $\hat{x}^{ij}$ is not assigned
  $\hat{x}^{ij} \leftarrow \min\{\bar{x}^{i*}, \bar{x}^{*j}\}$ $\{\hat{x}^{ij} \geq 0\}$
  $\bar{x}^{i*} \leftarrow \bar{x}^{i*} - \hat{x}^{ij}$ $\{\bar{x}^{i*} \geq 0\}$
  $\bar{x}^{*j} \leftarrow \bar{x}^{*j} - \hat{x}^{ij}$ $\{\bar{x}^{*j} \geq 0\}$
  $\{x_s^i = \sum_{j:(i,j) \text{ assigned}} \hat{x}^{ij} + \bar{x}^{i*}\}$
  $\{x_t^j = \sum_{i:(i,j) \text{ assigned}} \hat{x}^{ij} + \bar{x}^{*j}\}$
end while

The updates ensure that $\hat{x}^{ij}$, $\bar{x}^{i*}$ and $\bar{x}^{*j}$ stay non-negative and that the following modified marginalization constraints are still satisfied after each iteration:

$$\begin{aligned}\hat{x}^i_s &= \sum_{i,j:|i-j|<T} \hat{x}^{ij} + \sum_{j:|i-j|\geq T} \hat{x}^{ij} + \bar{x}^{i*} = \sum_j \hat{x}^{ij} + \bar{x}^{i*} \\ \hat{x}^j_t &= \sum_{i:|i-j|<T} \hat{x}^{ij} + \sum_{i:|i-j|\geq T} \hat{x}^{ij} + \bar{x}^{*j} = \sum_i \hat{x}^{ij} + \bar{x}^{*j}.\end{aligned}$$

We show that all $\bar{x}^{i*}$ and $\bar{x}^{*j}$ are 0 after termination of this algorithm. First, it cannot be that $\bar{x}^{i*} > 0$ and $\bar{x}^{*j} > 0$ for some i and j: if this is the case for $i, j : |i-j| < T$, we can increase $\hat{x}^{ij}$ and simultaneously strictly lowering the overall smoothness cost, thus contradicting that our initial solution was optimal. If $\bar{x}^{i*} > 0$ and $\bar{x}^{*j} > 0$ for some $i, j : |i-j| \geq T$, this contradicts the instructions ($\hat{x}^{ij} \leftarrow \min\{\bar{x}^{i*}, \bar{x}^{*j}\}$, $\bar{x}^{i*} \leftarrow \bar{x}^{i*} - \hat{x}^{ij}$, $\bar{x}^{*j} \leftarrow \bar{x}^{*j} - \hat{x}^{ij}$) in the algorithm above, which sets one of $\bar{x}^{i*}$ or $\bar{x}^{*j}$ to zero. W.l.o.g. some of the $\bar{x}^{i*}$ are strictly greater than 0, but all $\bar{x}^{*j}$ are 0. We have

$$1 = \sum_i \hat{x}^i_s = \sum_i \sum_j \hat{x}^{ij} + \bar{x}^{i*} = \sum_j \hat{x}^j_t + \bar{x}^{i*} = 1 + \bar{x}^{i*},$$

which is a contradiction. Hence all $\bar{x}^{i*}$ and $\bar{x}^{*j}$ have to be 0 at the end of the algorithm. We further have

$$\sum_{j:|i-j|\geq T} \hat{x}^{ij} = x^{i*} \qquad \text{and} \qquad \sum_{i:|i-j|\geq T} \hat{x}^{ij} = x^{*j}$$

and the pairwise smoothness costs are the same for x and $\hat{x}$ (similar to Eq. 19) and both overall objectives for $E_{\text{full}}(\hat{x})$ and $E_{\text{red}}(x)$ coincide. Thus, we have proved Observation 1.

6 Proof of Observation 2

We show that if we are given an optimal primal/dual solution pair generated by the refinement procedure satisfying the assumption stated in the observation, a primal-dual pair of optimality certificates can be constructed for the tight model, E_{tight}.

Note that the only difference between the dual of the tight model,

$$E^*_{\text{tight-I}}(p) = \sum_s \min_i \{\operatorname{div} p^i_s + \theta^i_s\} \qquad \text{s.t.} \qquad \|p^i_s - p^j_s\|_2 \leq \theta^{ij}, \tag{20}$$

and the weaker model for truncated costs,

$$\begin{aligned} E^*_{\text{fast}}(p) = &\sum_s \min_i \{\operatorname{div} p^i_s + \theta^i_s\} \\ \text{s.t. } &\|p^i_s - p^j_s\|_2 \leq \theta^{ij} && \forall s, \forall i, j : |i-j| < T \\ &\|p^i_s\| \leq \theta^*/2 && \forall s, i, \end{aligned} \tag{21}$$

is the set of constraints. We assume that $\theta^{ij} = \theta^*$ of $|i-j| > T$ in Eq. 20 and that $\theta^* \geq \theta^{ij}$, since we consider truncated smoothness cost. Consequently we have that the constraints in Eq. 21 are a superset of those in Eq. 20, due to $\|p^i_s\| \leq \theta^*/2$ implies $\|p^i_s - p^j_s\| \leq \theta^*$. The essential fact to prove observation 2 is, that if only two phase transitions are active, i.e. $y^{i_1*}_s \neq 0$ and $y^{i_2*}_s \neq 0$ for some i_1 and i_2, it must hold that $y^{i_1*}_s = -y^{i_2*}_s$ (the boundary normal of the entering phase must be opposite to the one of the leaving phase). This can be easily seen and is intuitive for the Potts smoothness cost. Extending that fact to general truncated smoothness priors can be seen as follows:

$$
\begin{aligned}
0 = \nabla \sum_i x_s^i = \sum_i \nabla x_s^i &= \sum_i \left(\sum_{j:i-T<j<i} y_s^{ji} - \sum_{j:i<j<i+T} y_s^{ij} - y_s^{i*} \right) \\
&= \sum_{i,j:i-T<j<i} y_s^{ji} - \sum_{i,j:i<j<i+T} y_s^{ij} - y_s^{i_1*} - y_s^{i_2*} \\
&= \sum_{i,j:i<j<i+T} y_s^{ij} - \sum_{i,j:i<j<i+T} y_s^{ij} - y_s^{i_1*} - y_s^{i_2*} \\
&= -y_s^{i_1*} - y_s^{i_2^*}.
\end{aligned}
$$

Note that from the normalization constraint, $\sum_i x_s^i = 1$, it follows that $\nabla \sum_i x_s^i = 0$. Further, by assumption we have $y_s^{i*} = 0$ for $i \neq i_1, i_2$. First order optimality conditions $y_s^{i*} \in \partial \imath\{\|-p_s^i\|_2 \le \theta^*/2\}$ (i.e. $y_s^{i_1*} \propto -p_s^{i_1}$ and $y_s^{i_2*} \propto -p_s^{i_2}$) imply that $p_s^{i_1} = -p_s^{i_2}$. Together with $\|p_s^{i_1}\| = \|p_s^{i_2}\| = \theta^*/2$ we obtain $\|p_s^{i_1} - p_s^{i_2}\| = \theta^*$.

In the following we assume $i_1 < i_2$ w.l.o.g. Given now the primal solution obtained from the refinement approach, we construct a feasible primal solution for the tight energy, i.e. we have to determine y_s^{ij} for $i,j : |i-j| \ge T$. We set in this case $y_s^{i_1 i_2} = y_s^{i_1*}$, and $y_s^{ij} = 0$ for $i,j : |i-j| \ge T$ otherwise. It can be easily checked that this choice for y_s^{ij} satisfies the marginalization constraints, i.e. one half of the optimality conditions. The dual variables p are a certificate for optimality, since $y_s^{i_1 i_2} \neq 0$ implies $\|p_s^{i_1} - p_s^{i_2}\| = \theta^*$ (i.e. the inequality constraint is tight), and for $i,j : |i-j| \ge T$ we have $y_s^{ij} = 0$ and $\|p_s^i - p_s^j\| \le \theta^*$. Overall, the other half of optimality conditions, $y_s^{ij} \neq 0 \implies \|p_s^i - p_s^j\| = \theta^{ij}$, and we have shown optimality of the constructed solution with respect to the tight energy E_{tight}.

7 Notes on smoothing-based optimization

7.1 A smooth version of $h^\theta(z) = \sqrt{2}\big[\|z\|_2 - \theta\big]_+$

By construction we know that the convex conjugate of h^θ is given by

$$(h^\theta)^*(x) = \theta \|x\|_2 + \imath\{\|x\| \le \sqrt{2}\}.$$

Thus, a smooth version of h^θ is the convex conjugate of

$$(h_\varepsilon^\theta)^*(x) = \theta \|x\|_2 + \imath\{\|x\|_2 \le \sqrt{2}\} + \frac{\varepsilon}{2}\|x\|_2^2.$$

Consequently,

$$h_\varepsilon^\theta(z) = \max_{x:\|x\|_2 \le \sqrt{2}} x^T z - \theta\|x\|_2 - \frac{\varepsilon}{2}\|x\|_2^2.$$

If we fix $\|x\|$, then an x colinear with z is maximizing the expression, hence we can reduce the problem by restricting x to be $x = cz$ for some $c \ge 0$. Hence, the above maximization problem is equivalent to

$$h_\varepsilon^\theta(z) = \max_{c \ge 0 : c\|z\|_2 \le \sqrt{2}} c\|z\|_2^2 - c\theta\|z\|_2 - \frac{\varepsilon}{2}c^2\|z\|_2^2.$$

We have $h_\varepsilon^\theta(0) = 0$, and in the following we assume $z \neq 0$, i.e. $\|z\|_2 > 0$. We have to analyze three cases:

- $c \in (0, \sqrt{2}/\|z\|_2)$: First order conditions on c yield

$$\|z\|_2^2 - \theta\|z\|_2 - \varepsilon c\|z\|_2^2 \stackrel{!}{=} 0$$

i.e.

$$c = \frac{\|z\|_2 - \theta}{\varepsilon\|z\|_2} \qquad \text{and} \qquad h_\varepsilon^\theta(z) = \frac{1}{2\varepsilon}\big(\|z\|_2 - \theta\big)^2$$

in this case. Note that $c > 0$ if $\|z\|_2 > \theta$.

- $c = 0$: This case is effective if $\|z\|_2 \le \theta$, and in this case we have

$$h_\varepsilon^\theta(z) = 0.$$

- $c = \sqrt{2}/\|z\|_2$: In this case we obtain

$$h_\varepsilon^\theta(z) = \sqrt{2}\big(\|z\|_2 - \theta\big) - \varepsilon.$$

This case is in effect if $c = \frac{\|z\|_2-\theta}{\varepsilon\|z\|_2} \ge \frac{\sqrt{2}}{\|z\|_2}$, i.e. $\|z\| \ge \theta + \sqrt{2}\varepsilon$.

Overall we obtain the smooth version of h^θ as stated in the main text.

7.2 Bound on the operator norm of A

To get the Lipschitz constant we again look at the A matrix and get an upper bound for $\|A\|_2$ via $\|A\|_2^2 \le \|A\|_1\|A\|_\infty$. Note that $\|A\|_1$ is the maximum absolute column sum, and $\|A\|_\infty$ is the maximum absolute row sum. The columns of A are indexed by the unknowns ($(p_s^i)_1$, $(p_s^i)_2$ and q_s), and the rows of A correspond to the terms in $E^*_{\text{tight-III}}$ (or its smooth version),

$$E^*_{\text{tight-III}}(p,q) = \sum_s q_s + \sum_{s,i}\big[\operatorname{div} p_s^i + \theta_s^i - q_s\big]_- + \sum_s \sum_{i,j:i<j} \sqrt{2}\min\big\{0, \theta^{ij} - \|p_s^i - p_s^j\|_2\big\}.$$

Since all occurrences of p_s^i and q_s have a $+1$ or -1 coefficient, it is sufficient to just count the occurrences of each variable. Since at most 5 variables appear in one term (rows corresponding to $[\operatorname{div} p_s^i + \theta_s^i - q_s]_-$), we have $\|A\|_\infty = 5$. q_s appears in $L+1$ terms (in q_s and in $\sum_i\big[\operatorname{div} p_s^i + \theta_s^i - q_s\big]_-$), and e.g. $(p_s^i)_1$ occurs also at most in $L+1$ terms (in the divergence terms with respect to s and $s-(1,0)$ and in $L-1$ expressions $\sum_{i,j:i<j}\sqrt{2}\min\big\{0,\theta^{ij} - \|p_s^i - p_s^j\|_2\big\}$), hence $\|A\|_1 = L+1$. Overall we have the bound $\|A\|_2^2 \le 5(L+1)$.

7.3 Extracting the primal solution from the smooth dual

We recall the smooth dual energy and indicate the correspondence between the terms in the dual energy and the respective primal variable,

$$-E^*_{\text{tight-III},\varepsilon}(p,q) = \sum_s -q_s + \sum_{s,i} \underbrace{\big[q_s - \operatorname{div} p_s^i - \theta_s^i\big]_{+,\varepsilon}}_{\triangleq x_s^i} + \sum_s \sum_{i,j:i<j} \underbrace{h_\varepsilon^{\theta^{ij}}(p_s^i - p_s^j)}_{\triangleq y_s^{ij}}. \tag{22}$$

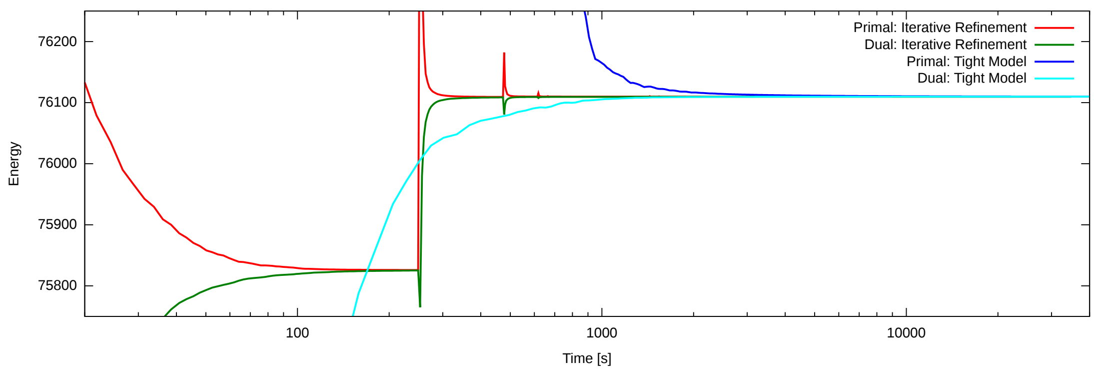
First order optimality conditions require that the corresponding primal unknowns are given by

$$x_s^i = \frac{d}{dz}\big[z - \theta_s^i\big]_{+,\varepsilon}\big|_{z=q_s-\operatorname{div} p_s^i}$$

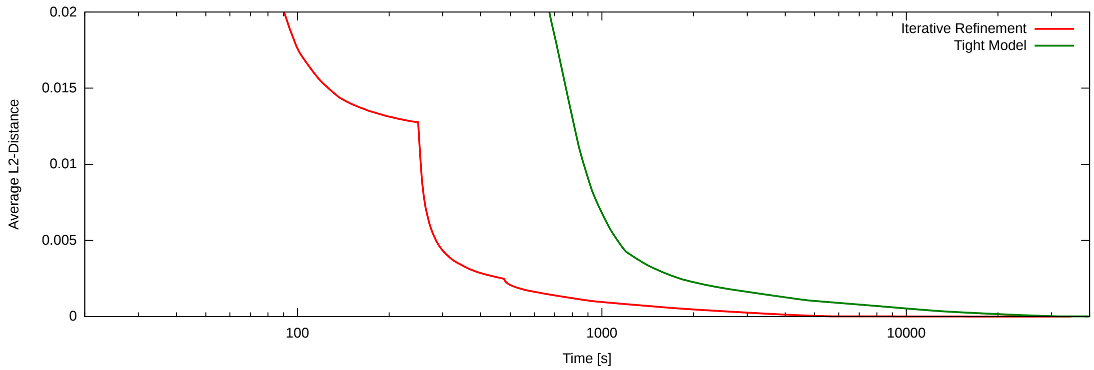
and

$$y_s^{ij} = \nabla_z h_\varepsilon^{\theta^{ij}}(z)\big|_{z=p_s^i-p_s^j}.$$

This allows to obtain primal estimates for iterative dual optimization methods, but the marginalization constraints between x_s and y_s will be only fulfilled after convergence.



(a) Primal and dual energy evolution



(b) Euclidean distance to a fully converged solution

Figure 1: Energy evolution and distance to the final solution for the Tsukuba stereo pair.

8 Numerical Convergence and Visual Comparison Between E_{tight} and E_{fast}

We use the standard Tsukuba stereo pair for illustration. The data term (unary potential) is

$$\lambda \sum_{c \in \{R,G,B\}} |I^c_{\text{left}}(x) - I^c_{\text{right}}(x + d)|$$

In Fig. 1 the evolution of the energies and of the distance to a converged solution is depicted (with $\lambda = 20$ and the Potts smoothness prior). The graphs are shown for direct optimization of the full model Eq. 3 and for the iterative refinement method (Section 4.1 in the main submission). Although there is very little difference in the visual results after a few 100 iterations, numerical convergence is slow (as usual for first-order methods applied on non-strict convex problems). Fig. 2 illustrates the visual difference between the tight and the efficient model for truncated linear smoothness costs. The values of λ are varying for the different truncation values in order to have roughly the same visual appearance. In real situations the difference between the tight and the efficient relaxations are smaller than for the triple junction inpainting example (due to the presence of the unary data term).

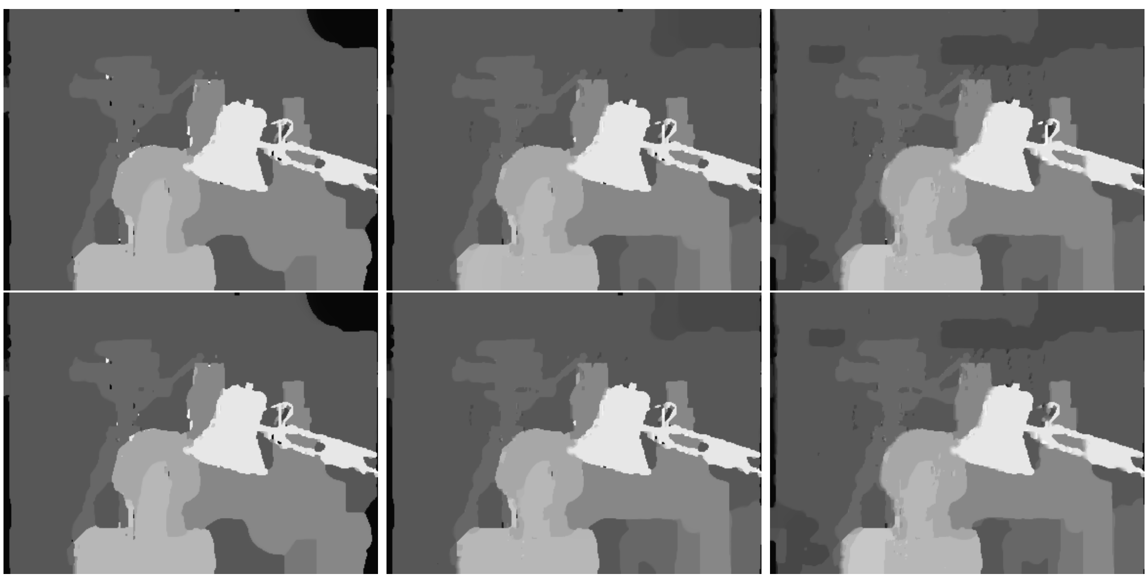


Figure 2: Visual comparison between the efficient and the tight relaxation. Top row: E_{fast}, bottom row: E_{tight}. 1st column: Potts model, $\lambda = 5$. 2nd column: truncated linear with truncation at 2, $\lambda = 10$. 3rd column: truncated linear with truncation at 4, $\lambda = 15$.