Continuous Relaxation of MAP Inference: A Nonconvex Perspective
Supplementary Material

D. Khuê Lê-Huu[1, 2] Nikos Paragios[1, 2, 3]
[1]CentraleSupélec, Université Paris-Saclay [2]Inria [3]TheraPanacea
{khue.le, nikos.paragios}@centralesupelec.fr

Abstract

We give proofs of the presented theoretical results in Appendix A, implementation details of the methods in Appendix B, and experiment details in Appendix C.

A. Proofs

A.1. Proof of Proposition 1

Clearly, BCD stops when there is no *strict* descent of the energy. Since the solution at each iteration is discrete and the number of nodes as well as the number of labels are finite, BCD must stop after a finite number of iterations. Suppose that this number is k: $E(\mathbf{x}^{(k+1)}) = E(\mathbf{x}^{(k)})$. At each inner iteration (*i.e.* Step 2 in Algorithm 1), the label of a node is changed to a new label only if the new label can produce *strictly* lower energy. Therefore, the labeling of $\mathbf{x}^{(k+1)}$ and $\mathbf{x}^{(k)}$ must be the same because they have the same energy, which implies $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$, *i.e.* $\mathbf{x}^{(k)}$ is a fixed point.

A.2. Proof of Equation (32)

Recall from (20) that

$$F(\mathbf{x}^1,\dots,\mathbf{x}^D) = \sum_{\alpha=1}^{D}\sum_{i_1\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1\dots i_\alpha} \bigotimes \left\{\mathbf{x}^1_{i_1},\dots,\mathbf{x}^\alpha_{i_\alpha}\right\}, \tag{43}$$

Clearly, the terms corresponding to any $\alpha < d$ do not involve $\mathbf{x}^d$. Thus, we can rewrite the above as

$$\begin{aligned} F(\mathbf{x}^1,\dots,\mathbf{x}^D) &= \mathrm{cst}(\mathbf{x}^{\mathrm{d}}) \\ &+ \sum_{\alpha=d}^{D}\sum_{i_1\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1\dots i_\alpha} \bigotimes \left\{\mathbf{x}^1_{i_1},\dots,\mathbf{x}^\alpha_{i_\alpha}\right\}. \end{aligned} \tag{44}$$

We will show that the last double sum can be written as $\sum_{i\in\mathcal{V}} \left\langle \mathbf{p}_i^d, \mathbf{x}_i^d\right\rangle$, where $\mathbf{p}_i^d$ is given by (32). The idea is to regroup, for each node i, all terms that contain $\mathbf{x}_i$. Indeed, for a given d we have the identity:

$$\sum_{i_1 i_2\dots i_\alpha\in\mathcal{C}} = \sum_{i_d\in\mathcal{V}}\ \sum_{i_1\dots i_{d-1} i_d i_{d+1}\dots i_\alpha\in\mathcal{C}}. \tag{45}$$

Therefore, the double sum in (44) becomes

$$\sum_{\alpha=d}^{D}\sum_{i_d\in\mathcal{V}}\ \sum_{i_1\dots i_{d-1} i_d i_{d+1}\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1\dots i_\alpha} \bigotimes \left\{\mathbf{x}^1_{i_1},\dots,\mathbf{x}^\alpha_{i_\alpha}\right\}. \tag{46}$$

Rearranging the first and second sums we obtain

$$\sum_{i_d\in\mathcal{V}}\sum_{\alpha=d}^{D}\ \sum_{i_1\dots i_{d-1} i_d i_{d+1}\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1\dots i_\alpha} \bigotimes \left\{\mathbf{x}^1_{i_1},\dots,\mathbf{x}^\alpha_{i_\alpha}\right\}. \tag{47}$$

With the change of variable $i \leftarrow i_d$ this becomes

$$\sum_{i\in\mathcal{V}}\sum_{\alpha=d}^{D}\ \sum_{i_1\dots i_{d-1} i i_{d+1}\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1\dots i_\alpha} \bigotimes \left\{\mathbf{x}^1_{i_1},\dots,\mathbf{x}^\alpha_{i_\alpha}\right\}. \tag{48}$$

Now by factoring out $\mathbf{x}_i^d$ for each $i \in \mathcal{V}$ the above becomes

$$\sum_{i\in\mathcal{V}}\left(\sum_{\alpha=d}^{D}\ \sum_{i_1\dots i_{d-1} i i_{d+1}\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1 i_2\dots i_\alpha} \bigotimes \left\{\mathbf{x}^1_{i_1},\dots,\mathbf{x}^{d-1}_{i_{d-1}},\mathbf{x}^{d+1}_{i_{d+1}},\dots,\mathbf{x}^\alpha_{i_\alpha}\right\}\right)^\top \mathbf{x}_i^d, \tag{49}$$

which is $\sum_{i\in\mathcal{V}} \left\langle \mathbf{p}_i^d, \mathbf{x}_i^d\right\rangle$, where $\mathbf{p}_i^d$ is given by (32), QED.

A.3. Proof of Equations (36)–(38)

See Appendix B.3, page 14 on the details of ADMM.

A.4. Proof of Proposition 2

For PGD and FW, the result holds for general continuously differentiable function $E(\cdot)$ and closed convex set $\mathcal{X}$.

We refer to [2] (Sections 2.2.2 and 2.3.2) for a proof. Below we give a proof for BCD.

In Proposition 1 we have shown that BCD reaches a discrete fixed point $\mathbf{x}^{(k)}$ after a finite number of iterations k. Now, we show that this fixed point is stationary. Define $\Delta_i = \left\{\mathbf{u} \in \mathbb{R}^{|\mathcal{S}_i|} : \mathbf{u} \geq \mathbf{0}, \mathbf{1}^\top \mathbf{u} = 1\right\}$ $\forall i \in \mathcal{V}$ and let $\mathbf{x}^* = \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$. At the last i^{th} inner iteration (11) we have:

$$E(\mathbf{x}_{[1,i-1]}^{(k+1)}, \mathbf{x}_i, \mathbf{x}_{[i+1,n]}^{(k)}) \geq E(\mathbf{x}_{[1,i-1]}^{(k+1)}, \mathbf{x}_i^{(k+1)}, \mathbf{x}_{[i+1,n]}^{(k)}) \quad (50)$$

for all $\mathbf{x}_i \in \Delta_i$, which is

$$E(\mathbf{x}_{[1,i-1]}^*, \mathbf{x}_i, \mathbf{x}_{[i+1,n]}^*) \geq E(\mathbf{x}_{[1,i-1]}^*, \mathbf{x}_i^*, \mathbf{x}_{[i+1,n]}^*) \quad (51)$$

for all $\mathbf{x}_i \in \Delta_i$. Define for each i the function

$$E_i^*(\mathbf{x}_i) = E(\mathbf{x}_1^*, \dots, \mathbf{x}_{i-1}^*, \mathbf{x}_i, \mathbf{x}_{i+1}^*, \dots, \mathbf{x}_n^*). \quad (52)$$

Obviously $E_i^*(\mathbf{x}_i)$ is continuously differentiable as it is linear. Since $\mathbf{x}_i^*$ is a minimizer of $E_i^*(\mathbf{x}_i)$ over Δ_i, which is closed and convex, according to (39) (which is a necessary optimality condition) we have $\nabla E_i^*(\mathbf{x}_i^*)^\top (\mathbf{x}_i - \mathbf{x}_i^*) \geq 0 \; \forall \mathbf{x}_i \in \Delta_i$. Notice that

$$\nabla E(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial E(\mathbf{x}^*)}{\partial \mathbf{x}_1} \\ \vdots \\ \frac{\partial E(\mathbf{x}^*)}{\partial \mathbf{x}_n} \end{bmatrix} = \begin{bmatrix} \nabla E_1^*(\mathbf{x}_1^*) \\ \vdots \\ \nabla E_n^*(\mathbf{x}_n^*) \end{bmatrix}, \quad (53)$$

we have

$$\nabla E(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) = \sum_{i=1}^n \nabla E_i^*(\mathbf{x}_i^*)^\top (\mathbf{x}_i - \mathbf{x}_i^*). \quad (54)$$

Since each term in the last sum is non-negative, we have $\nabla E(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \; \forall \mathbf{x} \in \mathcal{X}$, *i.e.* $\mathbf{x}^*$ is stationary.

A.5. Proof of Proposition 3

By Definition 2, a point $(\mathbf{x}^1, \dots, \mathbf{x}^D, \mathbf{y})$ is a KKT of (21) if and only if it has the form $(\mathbf{x}^*, \dots, \mathbf{x}^*, \mathbf{y}^*)$ (where $\mathbf{x}^* \in \mathcal{X}$) and at the same time satisfies

$$\mathbf{x}^{*d} \in \operatorname*{argmin}_{\mathbf{x}^d \in \mathcal{X}^d} \left\{F(\mathbf{x}^*, \dots, \mathbf{x}^*, \mathbf{x}^d, \mathbf{x}^*, \dots, \mathbf{x}^*) + \mathbf{y}^{*\top} \mathbf{A}^d \mathbf{x}^d\right\} \quad (55)$$

for all d, which is *equivalent* to

$$\left(\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^*, \dots, \mathbf{x}^*) + \mathbf{A}^{d\top} \mathbf{y}^*\right)^\top (\mathbf{x}^d - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x}^d \in \mathcal{X}^d, \forall d. \quad (56)$$

The equivalence ("⇔") follows from the fact that the objective function (with respect to $\mathbf{x}^d$) in (55) is convex. This is a well-known result in convex analysis, which we refer to Bertsekas, Dimitri P., Angelia Nedi, and Asuman E. Ozdaglar. *Convex analysis and optimization.*" (2003) (Proposition 4.7.2) for a proof. Note that from the necessary optimality condition (39) we can only have the "⇒" direction.

We need to prove that the sequence $\{(\mathbf{x}^{1^{(k)}}, \dots, \mathbf{x}^{D^{(k)}}, \mathbf{y}^{(k)})\}$ generated by ADMM satisfies the above conditions (under the assumption that the residual $r^{(k)}$ converges to 0).

Let $(\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*D}, \mathbf{y}^*)$ be a limit point of $\{(\mathbf{x}^{1^{(k)}}, \dots, \mathbf{x}^{D^{(k)}}, \mathbf{y}^{(k)})\}$ (thus $\mathbf{x}^{*d} \in \mathcal{X}^d \; \forall d$ since $(\mathcal{X}^d)_{1 \leq d \leq D}$ are closed), and define a subsequence that converges to this limit point by $\{(\mathbf{x}^{1^{(l)}}, \dots, \mathbf{x}^{D^{(l)}}, \mathbf{y}^{(l)})\}$, $l \in \mathcal{L} \subset \mathbb{N}$ where $\mathcal{L}$ denotes the set of indices of this subsequence. We have

$$\lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} (\mathbf{x}^{1^{(l)}}, \dots, \mathbf{x}^{D^{(l)}}, \mathbf{y}^{(l)}) = (\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*D}, \mathbf{y}^*). \quad (57)$$

Since the residual $r^{(k)}$ (30) converges to 0, we have

$$\lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} \left(\sum_{d=1}^D \mathbf{A}^d \mathbf{x}^{d^{(l)}}\right) = \mathbf{0}, \quad (58)$$

$$\lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} \left(\mathbf{x}^{d^{(l+1)}} - \mathbf{x}^{d^{(l)}}\right) = \mathbf{0} \quad \forall d. \quad (59)$$

On the one hand, combining (57) and (59) we get

$$\lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} (\mathbf{x}^{1^{(l+1)}}, \dots, \mathbf{x}^{D^{(l+1)}}, \mathbf{y}^{(l+1)}) = (\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*D}, \mathbf{y}^*). \quad (60)$$

(Note that the above is different from (57) because $l + 1$ might not belong to $\mathcal{L}$.) On the other hand, combining (57) and (58) we get

$$\sum_{d=1}^D \mathbf{A}^d \mathbf{x}^{*d} = \mathbf{0}, \quad (61)$$

which is, according to (22), equivalent to

$$\mathbf{x}^{*1} = \mathbf{x}^{*2} = \dots = \mathbf{x}^{*D}. \quad (62)$$

Let $\mathbf{x}^* \in \mathcal{X}$ denote the value of these vectors. From (57) and (60) we have

$$\lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} \mathbf{x}^{d^{(l)}} = \lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} \mathbf{x}^{d^{(l+1)}} = \mathbf{x}^* \quad \forall d, \quad (63)$$

$$\lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} \mathbf{y}^{(l)} = \lim_{\substack{l \to +\infty \\ l \in \mathcal{L}}} \mathbf{y}^{(l+1)} = \mathbf{y}^*. \quad (64)$$

It only remains to prove that $(\mathbf{x}^*, \dots, \mathbf{x}^*, \mathbf{y}^*)$ satisfies (56). Let us denote for convenience

$$\mathbf{z}_d^{(k)} = (\mathbf{x}^{[1,d]^{(k+1)}}, \mathbf{x}^{[d+1,D]^{(k)}}) \quad \forall d. \quad (65)$$

According to (39), the $\mathbf{x}$ update (28) implies

$$\left(\frac{\partial L_\rho}{\partial \mathbf{x}^d}(\mathbf{z}_d^{(k)},\mathbf{y}^{(k)})\right)^\top \left(\mathbf{x}^d - \mathbf{x}^{d^{(k+1)}}\right) \geq 0 \quad \forall \mathbf{x}^d \in \mathcal{X}^d, \forall d, \forall k. \quad (66)$$

Since L_ρ (27) is continuously differentiable, applying (63) and (64) we obtain

$$\lim_{\substack{l\to+\infty\\ l\in\mathcal{L}}} \frac{\partial L_\rho}{\partial \mathbf{x}^d}(\mathbf{z}_d^{(l)},\mathbf{y}^{(l)}) = \frac{\partial L_\rho}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*,\mathbf{y}^*) \quad \forall d. \quad (67)$$

Let $k=l$ in (66) and take the limit of that inequality, taking into account (63) and (67), we get

$$\left(\frac{\partial L_\rho}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*,\mathbf{y}^*)\right)^\top (\mathbf{x}^d - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x}^d \in \mathcal{X}^d, \forall d. \quad (68)$$

From the definition of L_ρ (27) we have

$$\begin{aligned}&\frac{\partial L_\rho}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*,\mathbf{y}^*)\\ =&\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*) + \mathbf{A}^{d\top}\mathbf{y}^* + \rho\mathbf{A}^{d\top}\left(\sum_{d=1}^D \mathbf{A}^d\mathbf{x}^*\right)\\ =&\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*) + \mathbf{A}^{d\top}\mathbf{y}^*. \quad (69)\end{aligned}$$

Note that the last equality follows from (22). Plugging the above into the last inequality we obtain

$$\left(\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*) + \mathbf{A}^{d\top}\mathbf{y}^*\right)^\top (\mathbf{x}^d - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x}^d \in \mathcal{X}^d, \forall d, \quad (70)$$

which is exactly (56), and this completes the proof.

A.6. Proof of Proposition 4

Let $(\mathbf{x}^*,\dots,\mathbf{x}^*,\mathbf{y}^*)$ be a KKT point of (21). We have seen in the previous proof that

$$\left(\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*) + \mathbf{A}^{d\top}\mathbf{y}^*\right)^\top (\mathbf{x}^d - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x}^d \in \mathcal{X}^d, \forall d. \quad (71)$$

According to (31):

$$\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^1,\dots,\mathbf{x}^D) = \mathbf{p}^d, \quad (72)$$

where $\mathbf{p}^d$ is defined by (32). Now let $\mathbf{p}^{*d}$ be the value of $\mathbf{p}^d$ where $(\mathbf{x}^1,\dots,\mathbf{x}^D)$ is replaced by $(\mathbf{x}^*,\dots,\mathbf{x}^*)$, *i.e.* $\mathbf{p}^{*d} = (\mathbf{p}_1^{*d},\dots,\mathbf{p}_n^{*d})$ where

$$\mathbf{p}_i^{*d} = \sum_{\alpha=d}^D \left(\sum_{i_1\dots i_{d-1} i i_{d+1}\dots i_\alpha\in\mathcal{C}} \mathbf{F}_{i_1 i_2\dots i_\alpha} \bigotimes \left\{\mathbf{x}_{i_1}^*,\dots,\mathbf{x}_{i_{d-1}}^*,\mathbf{x}_{i_{d+1}}^*,\dots,\mathbf{x}_{i_\alpha}^*\right\}\right) \forall i \in \mathcal{V}. \quad (73)$$

Notice that $\frac{\partial F}{\partial \mathbf{x}^d}(\mathbf{x}^*,\dots,\mathbf{x}^*) = \mathbf{p}^{*d}$, (71) becomes

$$\left(\mathbf{p}^{*d} + \mathbf{A}^{d\top}\mathbf{y}^*\right)^\top (\mathbf{x}^d - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x}^d \in \mathcal{X}^d, \forall d. \quad (74)$$

According to (23) we have $\mathcal{X} \subseteq \mathcal{X}^d$ and therefore the above inequality implies

$$\left(\mathbf{p}^{*d} + \mathbf{A}^{d\top}\mathbf{y}^*\right)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}, \forall d. \quad (75)$$

Summing this inequality for all d we get

$$\left(\sum_{d=1}^D \mathbf{p}^{*d}\right)^\top (\mathbf{x} - \mathbf{x}^*) + \mathbf{y}^{*\top}\left(\sum_{d=1}^D \mathbf{A}^d\right)(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}. \quad (76)$$

Yet, according to (22) we have $\sum_{d=1}^D \mathbf{A}^d\mathbf{x} = \sum_{d=1}^D \mathbf{A}^d\mathbf{x}^* = \mathbf{0}$. Therefore, the second term in the above inequality is 0, yielding

$$\left(\sum_{d=1}^D \mathbf{p}^{*d}\right)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}. \quad (77)$$

Now if we can prove that

$$\sum_{d=1}^D \mathbf{p}^{*d} = \nabla E(\mathbf{x}^*), \quad (78)$$

then we have $\nabla E(\mathbf{x}^*)^\top(\mathbf{x} - \mathbf{x}^*) \geq 0\ \forall \mathbf{x} \in \mathcal{X}$ and thus according to Definition 1, $\mathbf{x}^*$ is a stationary point of (**RLX**).

Let us now prove (78). Indeed, we can rewrite (73) as

$$\mathbf{p}_i^{*d} = \sum_{\alpha=d}^D \left(\sum_{\substack{C\in\mathcal{C}\\ C=(i_1\dots i_{d-1} i i_{d+1}\dots i_\alpha)}} \mathbf{F}_C \bigotimes \{\mathbf{x}_j^*\}_{j\in C\setminus i}\right) \quad \forall i \in \mathcal{V}. \quad (79)$$

Therefore,

$$\sum_{d=1}^D \mathbf{p}_i^{*d} = \sum_{d=1}^D \sum_{\alpha=d}^D \sum_{\substack{C\in\mathcal{C}\\ C=(i_1\dots i_{d-1} i i_{d+1}\dots i_\alpha)}} \mathbf{F}_C \bigotimes \{\mathbf{x}_j^*\}_{j\in C\setminus i} \quad \forall i \in \mathcal{V}. \quad (80)$$

Let's take a closer look at this triple sum. The double sum

$$\sum_{\alpha=d}^{D} \sum_{\substack{C\in\mathcal{C} \\ C=(i_1\dots i_{d-1} i i_{d+1}\dots i_\alpha)}}$$

basically means *iterating through all cliques whose sizes are* $\geq d$ *and whose* d^{th} *node is* i. Obviously the condition "sizes $\geq d$" is redundant here, thus the above means *iterating through all cliques whose* d^{th} *node is* i. Combined with $\sum_{d=1}^{D}$, the above triple sum means *for each size* d, *iterating through all cliques whose* d^{th} *node is* i, which is clearly equivalent to *iterating through all cliques that contain* i. Therefore, (80) can be rewritten more compactly as

$$\sum_{d=1}^{D} \mathbf{p}_i^{*d} = \sum_{C\in\mathcal{C}(i)} \mathbf{F}_C \bigotimes \{\mathbf{x}_j^*\}_{j\in C\setminus i} \quad \forall i \in \mathcal{V}, \tag{81}$$

where $\mathcal{C}(i)$ is the set of cliques that contain the node i. Recall from (14) that the last expression is actually $\frac{\partial E(\mathbf{x}^*)}{\partial \mathbf{x}^i}$, *i.e.*

$$\sum_{d=1}^{D} \mathbf{p}_i^{*d} = \frac{\partial E(\mathbf{x}^*)}{\partial \mathbf{x}^i} \quad \forall i \in \mathcal{V}, \tag{82}$$

or equivalently

$$\sum_{d=1}^{D} \mathbf{p}^{*d} = \nabla E(\mathbf{x}^*), \tag{83}$$

which is (78), and this completes the proof.

B. More details on the implemented methods

We present additional details on PGD, FW, ADMM as well as CQP (we omit BCD since it was presented with sufficient details in the paper).

Recall that our nonconvex relaxation is to minimize

$$E(\mathbf{x}) = \sum_{C\in\mathcal{C}} \mathbf{F}_C \bigotimes \{\mathbf{x}_i\}_{i\in C} \tag{10}$$

subject to $\mathbf{x} \in \mathcal{X} := \left\{ \mathbf{x} \;\middle|\; \mathbf{1}^\top \mathbf{x}_i = 1, \mathbf{x}_i \geq \mathbf{0} \; \forall i \in \mathcal{V} \right\}$.

B.1. PGD and FW

Recall from Section 4.1 that the main update steps in PGD and FW are respectively

$$\mathbf{s}^{(k)} = \operatorname*{argmin}_{\mathbf{s}\in\mathcal{X}} \left\| \mathbf{x}^{(k)} - \beta^{(k)} \nabla E(\mathbf{x}^{(k)}) - \mathbf{s} \right\|_2^2, \tag{84}$$

and

$$\mathbf{s}^{(k)} = \operatorname*{argmin}_{\mathbf{s}\in\mathcal{X}} \mathbf{s}^\top \nabla E(\mathbf{x}^{(k)}). \tag{85}$$

Clearly, in the PGD update step (84) the vector $\mathbf{s}^{(k)}$ is the projection of $\mathbf{x}^{(k)} - \beta^{(k)} \nabla E(\mathbf{x}^{(k)})$ onto $\mathcal{X}$. As we have discussed at the end of Section (4.2), this projection is reduced to independent projections onto the simplex $\left\{ \mathbf{x}_i \mid \mathbf{1}^\top \mathbf{x}_i = 1, \mathbf{x}_i \geq \mathbf{0} \right\}$ for each node i. In our implementation we used the method introduced in [6] for this simplex projection task.

The FW update step (85) can be solved independently for each node as well:

$$\mathbf{s}_i^{(k)} = \operatorname*{argmin}_{\mathbf{1}^\top \mathbf{s}_i = 1, \mathbf{s}_i \geq \mathbf{0}} \mathbf{s}_i^\top \frac{\partial E(\mathbf{x}^{(k)})}{\partial \mathbf{x}_i} \quad \forall i \in \mathcal{V}, \tag{86}$$

which is similar to the BCD update step (11) and thus can be solved using Lemma 1.

Next, we describe the line-search procedure (17) for these methods. Before going into details, we should note that in addition to line-search, we also implemented other step-size update rules such as *diminishing* or *Armijo* ones. However, we found that these rules do not work as well as line-search (the diminishing rule converges slowly while the search in the Armijo rule is expensive). We refer to [2] (Chapter 2) for further details on these rules.

Line search The line-search step consists of finding

$$\alpha^{(k)} = \operatorname*{argmin}_{0\leq\alpha\leq 1} E\left(\mathbf{x}^{(k)} + \alpha \mathbf{r}^{(k)}\right), \tag{87}$$

where $\mathbf{r}^{(k)} = \mathbf{s}^{(k)} - \mathbf{x}^{(k)}$. The term $E\left(\mathbf{x}^{(k)} + \alpha \mathbf{r}^{(k)}\right)$ is clearly a D^{th}-degree polynomial of α (recall that D is the degree of the MRF), which we denote $p(\alpha)$. If we can determine the coefficients of $p(\alpha)$, then (87) can be solved efficiently. In particular, if $D \leq 3$ then (87) has simple closed-form solutions (since the derivative of a 3rd-order polynomial is a 2nd-order one, which has simple closed-form solutions). For $D > 3$ we find that it is efficient enough to perform an exhaustive search over the interval $[0, 1]$ (with some increment value δ) for the best value of α. In the implementation we used $\delta = 0.0001$.

Now let us describe how to find the coefficients of $p(\alpha)$. For pairwise MRFs (*i.e.* $D = 2$), the energy is

$$E_{\text{pairwise}}(\mathbf{x}) = \sum_{i\in\mathcal{V}} \mathbf{F}_i^\top \mathbf{x}_i + \sum_{ij\in\mathcal{E}} \mathbf{x}_i^\top \mathbf{F}_{ij} \mathbf{x}_j, \tag{88}$$

where $\mathcal{E}$ is the set of edges, and thus

$$p(\alpha) = E_{\text{pairwise}}(\mathbf{x} + \alpha\mathbf{r}) \tag{89}$$

$$= \sum_{i\in\mathcal{V}} \mathbf{F}_i^\top (\mathbf{x}_i + \alpha\mathbf{r}_i) + \sum_{ij\in\mathcal{E}} (\mathbf{x}_i + \alpha\mathbf{r}_i)^\top \mathbf{F}_{ij} (\mathbf{x}_j + \alpha\mathbf{r}_j) \tag{90}$$

$$= A\alpha^2 + B\alpha + C, \tag{91}$$

where

$$A = \sum_{ij\in\mathcal{E}} \mathbf{r}_i^\top \mathbf{F}_{ij}\mathbf{r}_j \tag{92}$$

$$B = \sum_{i\in\mathcal{V}} \mathbf{F}_i^\top \mathbf{r}_i + \sum_{ij\in\mathcal{E}} \left(\mathbf{x}_i^\top \mathbf{F}_{ij}\mathbf{r}_j + \mathbf{r}_i^\top \mathbf{F}_{ij}\mathbf{x}_j\right) \tag{93}$$

$$C = E_{\text{pairwise}}(\mathbf{x}). \tag{94}$$

For higher-order MRFs, the analytical expressions of the polynomial coefficients are very complicated. Instead, we can find them numerically as follows. Since $p(\alpha)$ is a D^{th}-degree polynomial, it has $D+1$ coefficients, where the constant coefficient is already known:

$$p(0) = E(\mathbf{x}^{(k)}). \tag{95}$$

It remains D unknown coefficients, which can be computed if we have D equations. Indeed, if we evaluate $p(\alpha)$ at D different random values of α (which must be different than 0), then we obtain D linear equations whose variables are the coefficients of $p(\alpha)$. Solving this system of linear equations we get the values of these coefficients. This procedure requires D evaluations of the energy $E\left(\mathbf{x}^{(k)} + \alpha\mathbf{r}^{(k)}\right)$, but we find that it is efficient enough in practice.

B.2. Convex QP relaxation

This relaxation was presented in [19] for pairwise MRFs (88). Define:

$$d_i(s) = \sum_{j\in\mathcal{N}(i)} \sum_{t\in\mathcal{S}_j} \frac{1}{2}\left|f_{ij}(s,t)\right|. \tag{96}$$

Denote $\mathbf{d}_i = (d_i(s))_{s\in\mathcal{S}_i}$ and $\mathbf{D}_i = \operatorname{diag}(\mathbf{d}_i)$, the diagonal matrix composed by $\mathbf{d}_i$. The convex QP relaxation energy is given by

$$E_{\text{cqp}}(\mathbf{x}) = E_{\text{pairwise}}(\mathbf{x}) - \sum_{i\in\mathcal{V}} \mathbf{d}_i^\top \mathbf{x}_i + \sum_{i\in\mathcal{V}} \mathbf{x}_i^\top \mathbf{D}_i\mathbf{x}_i. \tag{97}$$

This convex energy can be minimized using different methods. Here we propose to solve it using Frank-Wolfe algorithm, which has the guarantee to reach the global optimum.

Similarly to the previous nonconvex Frank-Wolfe algorithm, the update step (85) can be solved using Lemma 1, and the line-search has closed-form solutions:

$$\begin{aligned} E_{\text{cqp}}(\mathbf{x}+\alpha\mathbf{r}) =& E_{\text{pairwise}}(\mathbf{x}+\alpha\mathbf{r}) - \sum_{i\in\mathcal{V}} \mathbf{d}_i^\top(\mathbf{x}_i+\alpha\mathbf{r}_i) \\ &+ \sum_{i\in\mathcal{V}} (\mathbf{x}_i+\alpha\mathbf{r}_i)^\top \mathbf{D}_i(\mathbf{x}_i+\alpha\mathbf{r}_i) \quad (98)\\ =& A'\alpha^2 + B'\alpha + C', \quad (99)\end{aligned}$$

where

$$A' = A + \sum_{i\in\mathcal{V}} \mathbf{r}_i^\top \mathbf{D}_i\mathbf{r}_i \tag{100}$$

$$B' = B + \sum_{i\in\mathcal{V}} \left(-\mathbf{d}_i^\top \mathbf{r}_i + \mathbf{r}_i^\top \mathbf{D}_i\mathbf{x}_i + \mathbf{x}_i^\top \mathbf{D}_i\mathbf{r}_i\right) \tag{101}$$

$$C' = C + \sum_{i\in\mathcal{V}} \left(-\mathbf{d}_i^\top \mathbf{x}_i + \mathbf{x}_i^\top \mathbf{D}_i\mathbf{x}_i\right). \tag{102}$$

B.3. ADMM

In this section, we give more details on the instantiation of ADMM into different decompositions. As we have seen in Section 4.2, there is an infinite number of such decompositions. Some examples include:

$$\begin{aligned} &\text{(cyclic)} && \mathbf{x}^{d-1} = \mathbf{x}^d, \quad d=2,\dots,D, && (103)\\ &\text{(star)} && \mathbf{x}^1 = \mathbf{x}^d, \quad d=2,\dots,D, && (104)\\ &\text{(symmetric)} && \mathbf{x}^d = (\mathbf{x}^1+\dots+\mathbf{x}^D)/D \quad \forall d. && (105)\end{aligned}$$

Let us consider for example the *cyclic* decomposition. We obtain the following problem, equivalent to (**RLX**):

$$\begin{aligned} \min \quad & F(\mathbf{x}^1,\mathbf{x}^2,\dots,\mathbf{x}^D) \\ \text{s.t.} \quad & \mathbf{x}^{d-1} = \mathbf{x}^d, \quad d=2,\dots,D, \\ & \mathbf{x}^d \in \mathcal{X}^d, \quad d=1,\dots,D, \end{aligned} \tag{106}$$

where $\mathcal{X}^1,\dots,\mathcal{X}^D$ are closed convex sets satisfying $\mathcal{X}^1\cap\mathcal{X}^2\cap\dots\cap\mathcal{X}^D = \mathcal{X}$, and F is defined by (20).

The augmented Lagrangian of this problem is:

$$\begin{aligned} &L_\rho(\mathbf{x}^1,\dots,\mathbf{x}^D,\mathbf{y}) = F(\mathbf{x}^1,\dots,\mathbf{x}^D) \\ &+ \sum_{d=2}^D \left\langle \mathbf{y}^d, \mathbf{x}^{d-1}-\mathbf{x}^d\right\rangle + \frac{\rho}{2}\sum_{d=2}^D \left\|\mathbf{x}^{d-1}-\mathbf{x}^d\right\|_2^2, \end{aligned} \tag{107}$$

where $\mathbf{y} = (\mathbf{y}^2,\dots,\mathbf{y}^D)$. The $\mathbf{y}$ update (29) becomes

$$\mathbf{y}^{d^{(k+1)}} = \mathbf{y}^{d^{(k)}} + \rho\left(\mathbf{x}^{d-1^{(k+1)}} - \mathbf{x}^{d^{(k+1)}}\right). \tag{108}$$

Consider the $\mathbf{x}$ update (28). Plugging (31) into (107), expanding and regrouping, we obtain that $L_\rho(\mathbf{x}^1,\dots,\mathbf{x}^D,\mathbf{y})$ is equal to each of the following expressions:

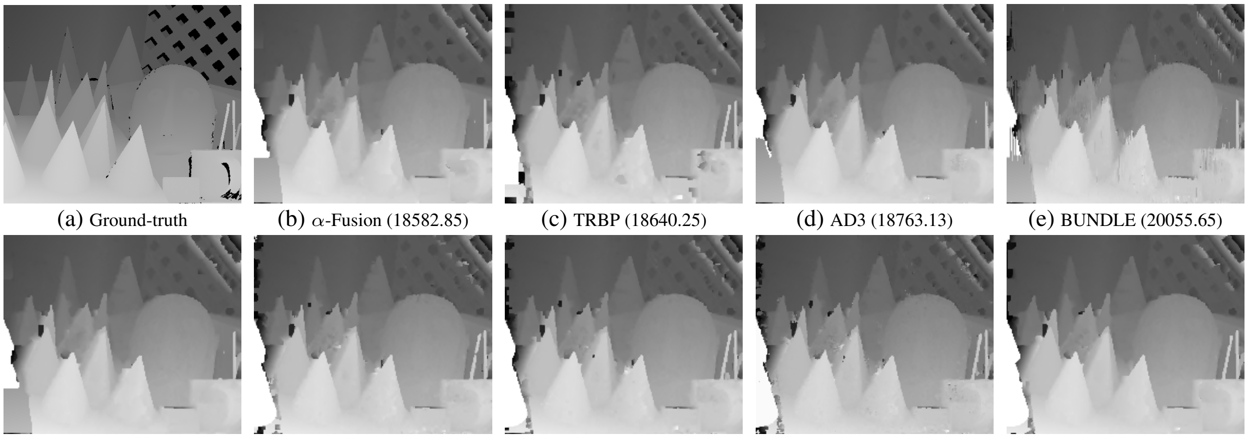
$$\frac{\rho}{2}\left\|\mathbf{x}^1\right\|_2^2 - \left\langle \mathbf{x}^1, \rho\mathbf{x}^2 - \mathbf{y}^2 - \mathbf{p}^1\right\rangle + \text{cst}(\mathbf{x}^1), \tag{109}$$

$$\begin{aligned}\rho\left\|\mathbf{x}^d\right\|_2^2 &- \left\langle \mathbf{x}^d, \rho\mathbf{x}^{d-1} + \rho\mathbf{x}^{d+1} + \mathbf{y}^d - \mathbf{y}^{d+1} - \mathbf{p}^d\right\rangle \\ &+ \text{cst}(\mathbf{x}^d) \qquad (2\le d\le D-1), \end{aligned} \tag{110}$$

$$\frac{\rho}{2}\left\|\mathbf{x}^D\right\|_2^2 - \left\langle \mathbf{x}^D, \rho\mathbf{x}^{D-1} + \mathbf{y}^D - \mathbf{p}^D\right\rangle + \text{cst}(\mathbf{x}^D). \tag{111}$$

From this, it is straightforward to see that the $\mathbf{x}$ update (28) is reduced to (35) where $(\mathbf{c}_d)_{1\le d\le D}$ are defined by (36), (37) and (38).

It is straightforward to obtain similar results for the other decompositions.



(a) Ground-truth (b) α-Fusion (18582.85) (c) TRBP (18640.25) (d) AD3 (18763.13) (e) BUNDLE (20055.65)

(f) SRMP (**18433.01**) (g) BCD (18926.70) (h) FW (18776.26) (i) PGD (19060.17) (j) ADMM (18590.87)

Figure 2: Resulted disparity maps and energy values using second-order MRFs for the *cones* scene of the Middlebury stereo dataset [21].

C. Details on the experiments

We replicated the model presented in [26] for the second-order stereo experiment, with some simplifications: we only used segmentation proposals (denoted by *SegPln* in [26]) and omitted the binary visibility variables and edges, so that all the nodes have the same number of labels. We ran the code provided by [26] to get the unary potentials as well as the 14 proposals, and then built the MRF model using OpenGM [1]. An example of resulted disparity maps for the *cones* scene of the Middlebury stereo dataset [21] is given in Figure 2.

For further details on the other modes, we refer to [10].

The detailed results of the experiments are provided at the end of this document.

Table 4: inpainting-n4

inpainting-n4		FastPD	α-Exp	TRBP	ADDD	MPLP	MPLP-C	TRWS	BUNDLE
triplepoint4-plain-ring-inverse	value	424.90	**424.12**	475.95	482.23	508.94	453.17	496.37	425.90
	bound	205.21	-Inf	-Inf	402.83	339.29	411.48	411.59	**411.87**
	runtime	0.03	0.02	33.04	28.59	1.75	3615.97	2.15	44.41
triplepoint4-plain-ring	value	**484.59**	**484.59**	**484.59**	**484.59**	485.38	**484.59**	**484.59**	**484.59**
	bound	384.57	-Inf	-Inf	**484.59**	484.58	484.59	**484.59**	484.59
	runtime	0.03	0.02	13.85	3.15	108.89	118.43	0.59	27.96
mean energy		454.75	454.35	480.27	483.41	497.16	468.88	467.70	455.25
mean bound		294.89	-Inf	-Inf	443.71	411.94	448.03	448.09	448.23
mean runtime		0.03	0.02	23.45	15.87	55.32	1867.20	1.37	36.18
best value		50.00	100.00	50.00	50.00	0.00	50.00	50.00	50.00
best bound		0.00	0.00	0.00	50.00	0.00	0.00	50.00	50.00
verified opt		0.00	0.00	0.00	50.00	0.00	0.00	50.00	0.00

Table 5: inpainting-n4

inpainting-n4		CQP	ADMM	BCD	FW	PGD
triplepoint4-plain-ring-inverse	value	2256.45	**424.12**	443.18	443.18	444.75
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	2.60	7.69	0.11	1.05	0.77
triplepoint4-plain-ring	value	542.57	**484.59**	528.57	533.29	534.86
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	1.24	12.00	0.11	1.15	0.85
mean energy		490.09	454.35	485.88	488.23	489.80
mean bound		-Inf	-Inf	-Inf	-Inf	-Inf
mean runtime		1.92	9.84	0.11	1.10	0.81
best value		0.00	100.00	0.00	0.00	0.00
best bound		0.00	0.00	0.00	0.00	0.00
verified opt		0.00	0.00	0.00	0.00	0.00

Table 6: inpainting-n8

inpainting-n8		α-Exp	FastPD	TRBP	ADDD	MPLP	MPLP-C	BUNDLE	TRWS
triplepoint4-plain-ring-inverse	value	434.84	434.84	496.40	714.42	442.42	463.88	435.32	504.97
	bound	-Inf	0.00	-Inf	406.71	412.37	413.49	**415.83**	413.20
	runtime	0.90	0.19	97.95	57.01	1107.98	3660.44	112.91	16.09
triplepoint4-plain-ring	value	**495.20**	**495.20**	**495.20**	495.85	495.52	**495.20**	**495.20**	**495.20**
	bound	-Inf	272.56	-Inf	495.18	494.72	**495.20**	495.04	494.71
	runtime	0.67	0.11	30.04	14.56	581.96	884.34	110.56	16.37
mean energy		465.02	465.02	494.02	605.14	468.83	469.78	465.26	466.80
mean bound		-Inf	136.28	-Inf	450.95	453.55	454.35	455.43	453.96
mean runtime		0.78	0.15	64.00	35.78	844.97	2272.39	111.74	16.23
best value		50.00	50.00	50.00	0.00	0.00	50.00	50.00	50.00
best bound		0.00	0.00	0.00	0.00	0.00	50.00	50.00	0.00
verified opt		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 7: inpainting-n8

inpainting-n8		CQP	ADMM	BCD	FW	PGD
triplepoint4-plain-ring-inverse	value	1819.57	**434.32**	438.95	446.19	446.19
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	20.78	38.71	0.31	6.33	6.55
triplepoint4-plain-ring	value	538.25	**495.20**	524.94	533.45	533.45
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	2.46	42.57	0.28	5.55	3.83
mean energy		489.82	464.76	481.95	489.82	489.82
mean bound		-Inf	-Inf	-Inf	-Inf	-Inf
mean runtime		11.62	40.64	0.29	5.94	5.19
best value		0.00	100.00	0.00	0.00	0.00
best bound		0.00	0.00	0.00	0.00	0.00
verified opt		0.00	0.00	0.00	0.00	0.00

Table 8: matching

matching		TRBP	ADDD	MPLP	MPLP-C	BUNDLE	TRWS	CQP	ADMM
matching0	value	60000000075.71	200000000047.27	90000000059.69	**19.36**	58.64	61.05	118.90	42.09
	bound	-Inf	11.56	10.96	**19.36**	11.27	11.02	-Inf	-Inf
	runtime	0.00	2.45	0.22	8.02	1.09	0.04	0.06	0.02
matching1	value	170000000090.50	70000000031.36	50000000030.34	**23.58**	10000000021.89	102.20	138.99	107.31
	bound	-Inf	20.13	18.47	**23.58**	17.48	18.52	-Inf	-Inf
	runtime	0.00	3.82	0.52	4.52	2.70	0.04	0.10	0.94

matching		TRBP	ADDD	MPLP	MPLP-C	BUNDLE	TRWS	CQP	ADMM
matching2	value	110000000096.00	20000000026.59	30000000025.18	**26.08**	20000000043.93	51.59	156.46	107.41
	bound	-Inf	22.97	21.07	**26.08**	19.87	21.18	-Inf	-Inf
	runtime	0.00	4.12	0.94	8.25	3.56	0.12	0.08	0.26
matching3	value	80000000066.03	130000000051.70	90000000051.81	**15.86**	10000000042.82	41.92	93.67	43.69
	bound	-Inf	10.72	10.15	**15.86**	9.25	10.14	-Inf	-Inf
	runtime	0.00	2.25	0.21	3.36	1.96	0.01	0.07	0.02
mean energy		97500000064.52	105000000039.23	65000000041.76	21.22	10000000041.82	63.52	127.01	75.12
mean bound		-Inf	16.35	15.16	21.22	14.47	15.22	-Inf	-Inf
mean runtime		0.00	3.16	0.47	6.04	2.33	0.05	0.08	0.31
best value		0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
best bound		0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
verified opt		0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00

Table 9: matching

matching		BCD	FW	PGD
matching0	value	43.61	56.10	49.45
	bound	-Inf	-Inf	-Inf
	runtime	0.00	0.19	8.08
matching1	value	118.00	77.31	79.01
	bound	-Inf	-Inf	-Inf
	runtime	0.00	23.66	21.36
matching2	value	139.74	89.46	62.40
	bound	-Inf	-Inf	-Inf
	runtime	0.00	55.74	19.28
matching3	value	38.09	43.98	43.21
	bound	-Inf	-Inf	-Inf
	runtime	0.00	0.81	4.11
mean energy		84.86	66.71	58.52
mean bound		-Inf	-Inf	-Inf
mean runtime		0.00	20.10	13.21
best value		0.00	0.00	0.00
best bound		0.00	0.00	0.00
verified opt		0.00	0.00	0.00

Table 10: mrf-stereo

mrf-stereo		FastPD	α-Exp	TRBP	ADDD	MPLP	MPLP-C	TRWS	BUNDLE
ted-gm	value	1344017.00	**1343176.00**	1460166.00	NaN	NaN	NaN	1346202.00	1563172.00
	bound	395613.00	-Inf	-Inf	NaN	NaN	NaN	**1337092.22**	1334223.01
	runtime	14.94	29.75	3616.74	NaN	NaN	NaN	391.34	3530.00
tsu-gm	value	370825.00	370255.00	411157.00	455874.00	369304.00	369865.00	369279.00	**369218.00**
	bound	31900.00	-Inf	-Inf	299780.16	367001.47	366988.29	369217.58	**369218.00**
	runtime	1.72	3.64	1985.50	1066.79	4781.02	4212.26	393.76	670.81
ven-gm	value	3127923.00	3138157.00	3122190.00	NaN	NaN	NaN	**3048404.00**	3061733.00
	bound	475665.00	-Inf	-Inf	NaN	NaN	NaN	**3047929.95**	3047785.37
	runtime	4.76	10.87	2030.13	NaN	NaN	NaN	478.49	1917.58
mean energy		1614255.00	1617196.00	1664504.33	NaN	NaN	NaN	1587596.67	1664707.67
mean bound		301059.33	-Inf	-Inf	NaN	NaN	NaN	1584746.58	1583742.13
mean runtime		7.14	14.75	2544.12	NaN	NaN	NaN	421.20	2039.47
best value		0.00	33.33	0.00	0.00	0.00	0.00	33.33	33.33
best bound		0.00	0.00	0.00	0.00	0.00	0.00	66.67	33.33
verified opt		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 11: mrf-stereo

mrf-stereo		CQP	ADMM	BCD	FW	PGD
ted-gm	value	4195611.00	1373030.00	3436281.00	3020579.00	2694493.00
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	3602.97	3628.80	15.64	1740.10	2109.65
tsu-gm	value	3621062.00	375954.00	2722934.00	2352499.00	2114223.00
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	3600.79	807.70	5.33	622.64	120.38
ven-gm	value	26408665.00	3123334.00	14907352.00	13114176.00	10818561.00
	bound	-Inf	-Inf	-Inf	-Inf	-Inf
	runtime	3602.28	2696.49	11.48	3604.63	2298.42
mean energy		11408446.00	1624106.00	7022189.00	6162418.00	5209092.33
mean bound		-Inf	-Inf	-Inf	-Inf	-Inf
mean runtime		3602.01	2377.66	10.82	1989.12	1509.49
best value		0.00	0.00	0.00	0.00	0.00
best bound		0.00	0.00	0.00	0.00	0.00
verified opt		0.00	0.00	0.00	0.00	0.00

Table 12: inclusion

inclusion		α-Fusion	TRBP	ADDD	MPLP	MPLP-C	BUNDLE	SRMP	ADMM
modelH-1-0.8-0.2	value	1595.06	1416.07	2416.58	3416.08	5415.89	5427.91	**1415.94**	**1415.94**
	bound	-Inf	-Inf	1415.71	1415.70	1415.71	1406.09	**1415.94**	-Inf
	runtime	0.06	21.93	10.52	12.17	3843.79	99.77	0.11	106.70
modelH-10-0.8-0.2	value	1590.97	1416.80	3415.92	5415.13	4415.43	5422.47	**1416.10**	1416.24
	bound	-Inf	-Inf	1415.68	1415.62	1415.70	1404.47	**1416.10**	-Inf
	runtime	0.05	22.66	1.20	12.03	3797.40	91.16	0.13	85.46
modelH-2-0.8-0.2	value	1603.85	1423.42	4423.49	6422.84	3423.03	5436.16	**1422.89**	**1422.89**
	bound	-Inf	-Inf	1422.79	1422.78	1422.79	1411.56	**1422.89**	-Inf
	runtime	0.05	21.34	10.00	6.67	4051.20	101.83	0.11	113.24
modelH-3-0.8-0.2	value	1596.11	**1381.14**	**1381.14**	**1381.14**	**1381.14**	4389.78	**1381.14**	1381.19
	bound	-Inf	-Inf	**1381.14**	1381.14	1381.14	1371.29	**1381.14**	-Inf
	runtime	0.06	8.02	4.50	7.79	8.84	112.52	0.11	63.51
modelH-4-0.8-0.2	value	1595.12	1427.56	5427.63	5426.48	3427.27	2432.97	**1427.17**	**1427.17**
	bound	-Inf	-Inf	1426.58	1426.56	1426.58	1416.80	**1427.17**	-Inf
	runtime	0.04	21.18	9.40	8.29	3892.65	116.38	0.13	125.01
modelH-5-0.8-0.2	value	1566.58	3383.89	6383.61	4383.52	6382.77	4390.47	**1383.69**	1383.77
	bound	-Inf	-Inf	1383.25	1383.23	1383.30	1371.94	**1383.69**	-Inf
	runtime	0.04	21.05	8.45	5.44	3902.54	112.86	0.18	99.08
modelH-6-0.8-0.2	value	1588.33	2402.30	2402.17	2402.60	5401.70	3406.27	**1402.34**	1402.60
	bound	-Inf	-Inf	1402.01	1401.77	1402.01	1393.05	**1402.34**	-Inf
	runtime	0.03	20.80	2.69	22.61	3778.21	101.74	0.11	126.40
modelH-7-0.8-0.2	value	1583.36	1403.61	3403.70	5402.97	5403.24	6418.08	**1403.25**	1403.69
	bound	-Inf	-Inf	1403.08	1403.07	1403.08	1391.87	**1403.25**	-Inf
	runtime	0.04	20.80	2.50	11.98	4124.95	103.95	0.15	94.36
modelH-8-0.8-0.2	value	1574.64	3368.65	3368.65	3368.66	1368.55	**1368.33**	**1368.33**	**1368.33**
	bound	-Inf	-Inf	1368.29	1368.29	1368.33	1368.23	**1368.33**	-Inf
	runtime	0.05	20.66	11.21	5.09	3740.80	92.39	0.15	86.69
modelH-9-0.8-0.2	value	1577.25	1385.00	1385.23	2385.04	3385.06	**1384.86**	**1384.86**	1384.95
	bound	-Inf	-Inf	1384.82	1384.82	1384.82	1384.81	**1384.86**	-Inf
	runtime	0.03	3.61	3.15	4.75	3824.62	82.98	0.11	73.29
mean energy		1587.13	1441.43	1694.72	3300.67	2800.54	4007.73	1400.57	1400.68
mean bound		-Inf	-Inf	1400.33	1400.30	1400.35	1392.01	1400.57	-Inf
mean runtime		0.05	18.20	6.36	9.68	3496.50	101.56	0.13	97.37
best value		0.00	10.00	10.00	10.00	10.00	20.00	100.00	40.00
best bound		0.00	0.00	10.00	0.00	0.00	0.00	100.00	0.00
verified opt		0.00	0.00	10.00	0.00	0.00	0.00	100.00	0.00

Table 13: inclusion

inclusion		BCD	FW	PGD
modelH-1-0.8-0.2	value	12435.37	7419.38	7421.24
	bound	-Inf	-Inf	-Inf
	runtime	0.14	44.22	67.47
modelH-10-0.8-0.2	value	15446.57	7427.81	5424.26
	bound	-Inf	-Inf	-Inf
	runtime	0.14	2.76	16.90
modelH-2-0.8-0.2	value	10430.00	5425.92	5425.74
	bound	-Inf	-Inf	-Inf
	runtime	0.14	11.55	57.53
modelH-3-0.8-0.2	value	15397.00	1382.80	1382.23
	bound	-Inf	-Inf	-Inf
	runtime	0.14	20.57	19.35
modelH-4-0.8-0.2	value	15447.30	4427.73	4427.66
	bound	-Inf	-Inf	-Inf
	runtime	0.13	8.25	109.47
modelH-5-0.8-0.2	value	9391.02	6385.98	6385.44
	bound	-Inf	-Inf	-Inf
	runtime	0.13	6.26	32.41
modelH-6-0.8-0.2	value	13420.27	5407.69	3403.83
	bound	-Inf	-Inf	-Inf
	runtime	0.14	36.05	24.21
modelH-7-0.8-0.2	value	11438.71	10411.17	11498.09
	bound	-Inf	-Inf	-Inf
	runtime	0.13	18.45	72.97
modelH-8-0.8-0.2	value	14385.72	6376.91	6375.75
	bound	-Inf	-Inf	-Inf
	runtime	0.14	35.24	80.66
modelH-9-0.8-0.2	value	7393.92	3386.31	3385.93
	bound	-Inf	-Inf	-Inf
	runtime	0.14	28.90	29.45
mean energy		12518.59	5805.17	5513.02
mean bound		-Inf	-Inf	-Inf
mean runtime		0.14	21.23	51.04
best value		0.00	0.00	0.00
best bound		0.00	0.00	0.00

inclusion		BCD	FW	PGD
verified opt		0.00	0.00	0.00

Table 14: stereo

stereo		α-Fusion	TRBP	ADDD	MPLP	MPLP-C	BUNDLE	SRMP	ADMM
art_small	value	13262.49	13336.35	13543.70	NaN	NaN	15105.28	**13091.20**	13297.79
	bound	-Inf	-Inf	12925.76	NaN	NaN	12178.62	**13069.30**	-Inf
	runtime	50.99	3744.91	3096.10	NaN	NaN	3845.89	3603.89	3710.92
cones_small	value	18582.85	18640.25	18763.13	NaN	NaN	20055.65	**18433.01**	18590.87
	bound	-Inf	-Inf	18334.00	NaN	NaN	17724.56	**18414.29**	-Inf
	runtime	48.89	3660.77	7506.15	NaN	NaN	3814.74	3603.11	3659.15
teddy_small	value	14653.53	14680.21	14804.46	NaN	NaN	15733.15	**14528.74**	14715.83
	bound	-Inf	-Inf	14374.12	NaN	NaN	13981.71	**14518.03**	-Inf
	runtime	50.99	3670.35	3535.79	NaN	NaN	3820.05	3603.49	3620.84
venus_small	value	9644.78	9692.80	9796.44	NaN	NaN	9990.68	**9606.34**	9669.62
	bound	-Inf	-Inf	9377.05	NaN	NaN	9402.97	**9601.86**	-Inf
	runtime	49.24	3627.58	3761.29	NaN	NaN	3774.66	3603.14	3657.60
mean energy		14035.91	14087.40	14226.93	NaN	NaN	15221.19	13914.82	14068.53
mean bound		-Inf	-Inf	13752.73	NaN	NaN	13321.96	13900.87	-Inf
mean runtime		50.03	3675.90	4474.83	NaN	NaN	3813.84	3603.41	3662.13
best value		0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00
best bound		0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00
verified opt		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 15: stereo

stereo		BCD	FW	PGD
art_small	value	13896.67	13696.50	13929.06
	bound	-Inf	-Inf	-Inf
	runtime	60.63	1407.07	3648.00
cones_small	value	18926.70	18776.26	19060.17
	bound	-Inf	-Inf	-Inf
	runtime	57.40	2111.63	3669.24
teddy_small	value	14998.31	14891.12	15193.23
	bound	-Inf	-Inf	-Inf
	runtime	60.08	1626.66	3671.59
venus_small	value	9767.21	9726.27	9992.13
	bound	-Inf	-Inf	-Inf
	runtime	60.27	1851.40	3670.82
mean energy		14397.22	14272.54	14543.65
mean bound		-Inf	-Inf	-Inf
mean runtime		59.59	1749.19	3664.92
best value		0.00	0.00	0.00
best bound		0.00	0.00	0.00
verified opt		0.00	0.00	0.00