A. Proofs

The BLP relaxation [17] introduces a probability distribution μ_i over $\{0,1\}$ for each $i \in [d]$ and a probability distribution μ_t over $\operatorname{dom} f_t$ for each $t \in T$. It can be written as follows:

$$\begin{array}{lll} \min\limits_{\mu \geq \mathbf{0}} & \sum\limits_{t \in T} \sum\limits_{z \in \operatorname{dom} f_t} \mu_t(z) f_t(z) & \\ \text{s.t.} & \mu_i(0) + \mu_i(1) = 1 & \forall i \in [d] \\ & \sum\limits_{z \in \operatorname{dom} f_t} \mu_t(z) = 1 & \forall t \in T \\ & \sum\limits_{z \in \operatorname{dom} f_t : z_i = a} \mu_t(z)_i = \mu_i(a) & \forall t \in T, i \in A_t, a \in \{0,1\} \end{array} \tag{19}$$

Let us show that the optimal values of (19) and (3) coincide.

Proof of equivalence of (19) *and* (3). Define an extension $\hat{f}_t : \mathbb{R}^{A_t} \to \mathbb{R} \cup \{+\infty\}$ of function $\hat{f} : \{0,1\}^{A_t} \to \mathbb{R} \cup \{+\infty\}$ as follows: for a vector $x \in \mathbb{R}^{A_t}$ set

$$\begin{array}{rl} \hat{f}_t(x) = \min\limits_{\mu_t \geq \mathbf{0}} & \sum\limits_{z \in \operatorname{dom} f_t} \mu_t(z) f_t(z) \\ \text{s.t.} & \sum\limits_{z \in \operatorname{dom} f_t} \mu_t(z) = 1 \\ & \sum\limits_{z \in \operatorname{dom} f_t} \mu_t(z) \cdot z = x \end{array} \tag{20}$$

Note, if $x \notin [0,1]^{A_t}$ then (20) does not have a feasible solution, and so $\hat{f}_t(x) = +\infty$. Observe that the constraints in the last line of (19) for $a = 0$ are redundant - they follow from the remaining constraints. Also observe that constraints $\sum_{z \in \operatorname{dom} f_t : z_i = 1} \mu_t(z)_i = \mu_i(1)$ for $i \in A_t$ can be written as $\sum_{z \in \operatorname{dom} f_t} \mu_t(z) \cdot z = x$ if we denote $x_i = \mu_i(1)$ for $i \in A_t$. Therefore, problem (19) can be equivalently rewritten as follows:

$$\min_{x \in \mathbb{R}^d} \sum_{t \in T} \hat{f}(x_{A_t}) \tag{21}$$

It can be seen that the last problem is equivalent to (3). Indeed, we just need to observe that for each $t \in T$ and $x \in \mathbb{R}^{A_t}$ we have

$$\min_{\substack{y \in \operatorname{conv}(\mathcal{Y}_t) \\ y_* = x}} y_\circ = \min_{\substack{\alpha \geq \mathbf{0},\ \sum_{z \in \operatorname{dom} f_t} \alpha(z) = 1 \\ y = \sum_{z \in \operatorname{dom} f_t} \alpha(z) \cdot [z\ f(z)] \\ y_* = x}} y_\circ$$

$$= \min_{\substack{\alpha \geq \mathbf{0},\ \sum_{z \in \operatorname{dom} f_t} \alpha(z) = 1 \\ \sum_{z \in \operatorname{dom} f_t} \alpha(z) \cdot z = x}} \sum\nolimits_{z \in \operatorname{dom} f_t} \alpha(z) f(z) = \hat{f}_t(x)$$

□

Proof of Proposition 1. Write $f(y) := \sum_{t \in T} y^t_\circ$, then problem (3) can be written as

$$\min_{\substack{(y,x) \in \mathbb{Y} \times \mathbb{R}^d \\ y^t_\star = x_{A_t}\ \forall t \in T}} f(y) \tag{22}$$

The Lagrangian w.r.t. the equality constraints is given by

$$\begin{aligned} L(y,x,\lambda) &= f(y) + \sum_{t \in T} \langle y^t_\star - x_{A_t}, \lambda^t \rangle \\ &= \sum_{t \in T} \langle y^t, [\lambda^t\ 1] \rangle - \sum_{t \in T} \langle x_{A_t}, \lambda^t \rangle \end{aligned}$$

Therefore, the dual function for $\lambda \in \bigotimes_{t \in T} \mathbb{R}^{A_t}$ is

$$\begin{aligned} h(\lambda) &= \min_{(y,x) \in \mathbb{Y} \times \mathbb{R}^d} L(y,x,\lambda) \\ &= \begin{cases} \sum\limits_{t \in T} \min\limits_{y^t \in \mathbb{Y}_t} \langle y^t, [\lambda^t\ 1] \rangle & \text{if } \lambda \in \Lambda \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

The problem can thus be formulated as $\max_\lambda h(\lambda)$, or equivalently as $\max_{\lambda \in \Lambda} h(\lambda)$. This coincides with formulation given in Proposition 1.

Since constraint $y \in \mathbb{Y}$ can be expressed as a linear program, the duality between (3) and (5) can be viewed as a special case of linear programming (LP) duality (where the value of function $h(\lambda)$ is also written as a resulting of some LP). For LPs it is known that strong duality holds assuming that either the primal or the dual problems have a feasible solution. This holds in our case, since vector $\lambda = \mathbf{0} \in \Lambda$ is feasible. We can conclude that we have a strong duality between (3) and (5). □

Proof of Proposition 2. First, we derive the dual of $h_{\mu,c}$:

$$\begin{aligned} & \max_{\lambda \in \Lambda} h_{\mu,c}(\lambda) \\ = & \max_{\lambda \in \Lambda} \sum_{t \in T} \min_{y^t \in \mathbb{Y}_t} \langle y^t, [\lambda^t\ 1] \rangle - \tfrac{1}{2c} \|\lambda^t - \mu^t\|^2 \\ = & \min_{y \in \mathbb{Y}} \underbrace{\max_{\lambda \in \Lambda} \sum_{t \in T} \langle y^t, [\lambda^t\ 1] \rangle - \frac{1}{2c} \|\lambda^t - \mu^t\|^2}_{=: f_{\mu,c}(y)} \end{aligned}$$

The function $f_{\mu,c}(y)$ has a closed form expression, since it is a quadratic function subject to linear equalities. Write $\nu_i = \frac{1}{|T_i|} \sum_{t \in T_i} (c \cdot y^t_i + \mu^t_i)$ for $i \in [d]$. The $\arg\max$ in the expression defining $f_{\mu,c}(y)$ are

$$\lambda^t = (c \cdot y^t_\star + \mu^t) - \nu_{A_t} \tag{23}$$

The function value is

$$\begin{aligned} f_{\mu,c}(x) &= \sum_{t \in T} \langle y^t, [\lambda^t_\star\ 1] \rangle - \tfrac{1}{2c} \|\lambda^t_\star - \mu^t\|^2 \\ &= \sum_{t \in T} \begin{pmatrix} \langle y^t_\star, c \cdot y^t_\star + \mu^t - \nu_{A_t} \rangle + y^t_\circ \\ -\tfrac{1}{2c} \|c x^t_\star + \mu^t - \nu_{A_t} - \mu^t\|^2 \end{pmatrix} \\ &= \sum_{t \in T} \begin{pmatrix} c\|y^t_\star\|^2 + \langle y^t_\star, \mu^t - \nu_{A_t} \rangle + y^t_\circ \\ -\tfrac{1}{2c} \|c y^t_\star - \nu_{A_t}\|^2 \end{pmatrix} \\ &= \sum_{t \in T} \begin{pmatrix} c\|y^t_\star\|^2 + \langle y^t_\star, \mu^t - \nu_{A_t} \rangle + y^t_\circ \\ -\tfrac{1}{2c} \left\{ \|c y^t_\star\|^2 - 2c \langle y^t_\star, \nu_{A_t} \rangle + \|\nu_{A_t}\|^2 \right\} \end{pmatrix} \\ &= \sum_{t \in T} \left(\tfrac{c}{2} \|y^t_\star\|^2 + \langle y^t_\star, \mu^t \rangle + y^t_\circ - \tfrac{1}{2c} \|\nu_{A_t}\|^2 \right). \end{aligned}$$

The gradient is $\nabla_t f_{\mu,c}(y) = [c \cdot y^t + \mu^t - \nu_{A_t}\ 1] = [\lambda^t\ 1]$. ☐

Proof of Proposition 3. Let $\overline{\mathbb{Y} \times \mathbb{R}^d}$ be the set of vectors $(y, x) \in \mathbb{Y} \times \mathbb{R}^d$ satisfying the equality constraints $y^t_\star = x_{A_t}$ for all t. By construction, for any $\lambda \in \Lambda$ we have

$$f(y) = L(y, x, \lambda) \quad \forall (y, x) \in \overline{\mathbb{Y} \times \mathbb{R}^d} \tag{24a}$$
$$L(y, x, \lambda) \geq h(\lambda) \quad \forall (y, x) \in \mathbb{Y} \times \mathbb{R}^d \tag{24b}$$
$$L(y, x, \lambda) = \sum_{t \in T} \langle y^t, [\lambda^t\ 1] \rangle \tag{24c}$$

Eq. (24c) gives that $A_{y,\lambda} = L(y, x, \lambda) - h(y)$ for any $(y, \lambda) \in \mathbb{Y} \times \Lambda$ and $x \in \mathbb{R}^d$, and so from (24b) we get that $A_{y,\lambda} \geq 0$. Clearly, we have $B_y \geq 0$. The following two facts imply part (b) of Proposition 3:

- Consider vector $y \in \mathbb{Y}$. Then $B_y = 0$ if and only if $(y, x) \in \overline{\mathbb{Y} \times \mathbb{R}^d}$ for some x. (This can be seen from the definition of B_y in Section 2.3).

- Consider vectors $(y, x) \in \overline{\mathbb{Y} \times \mathbb{R}^d}$ and $\lambda \in \Lambda$. They are an optimal primal-dual pair if and only if $f(y) = h(\lambda)$, which in turn holds if and only if $A_{y,\lambda} = 0$ (since $A_{y,\lambda} = L(y, x, \lambda) - h(\lambda) = f(y) - h(\lambda)$).

It remains to show inequality (11). Denote $\delta = \lambda^* - \lambda$, then $\sum_{t \in T_i} \delta^t_i = 0$ for any $i \in [d]$. Denoting $y^-_i = \min_{t \in T_i} y^t_i$ and $y^+_i = \max_{t \in T_i} y^t_i$, we get

$$\begin{aligned} \sum_{t \in T_t} y^t_i \cdot \delta^t_i &= \sum_{t \in T_t} \left[y^t_i - y^-_i \right] \cdot \delta^t_i \\ &\leq \sum_{t \in T_t} \left[y^+_i - y^-_i \right] \cdot |\delta^t_i| \\ &\leq \left[y^+_i - y^-_i \right] \cdot \|\delta\|_{1,\infty} \end{aligned}$$

Summing these inequalities over $i \in [d]$ gives

$$\sum_{t \in T} \langle y^t_\star, \delta^t \rangle \leq B_y \cdot \|\delta\|_{1,\infty}$$

Recalling that $A_{\lambda^*,y} \geq 0$, we obtain the desired claim:

$$\begin{aligned} h(\lambda^*) &\leq \sum_{t \in T} \langle y^t, [(\lambda^*)^t\ 1] \rangle \\ &= \sum_{t \in T} \langle y^t, [\lambda^t\ 1] \rangle + \sum_{t \in T} \langle y^t_\star, \delta^t \rangle \\ &\leq \sum_{t \in T} \langle y^t, [\lambda^t\ 1] \rangle + B_y \cdot \|\delta\|_{1,\infty} \end{aligned}$$

☐

Lemma 1 (step size in Algorithm 1)**.** *The optimal step size γ in Algorithm 1 is*

$$\gamma = \frac{\langle \nabla_t f_{\mu,c}(y), y^t - z^t \rangle}{c \|y^t_\star - z^t_\star\|^2} = \frac{\langle [c \cdot y^t_\star + \mu^t - \nu_{A_t}\ 1], y^t - z^t \rangle}{c \|y^t_\star - z^t_\star\|^2} \tag{25}$$

and clip γ to [0,1].

Proof. Recall that $y(\gamma)$ in algorithm 1 is defined as $y(\gamma)^s = \begin{cases} y^s, & s \neq t \\ (1-\gamma) y^t + \gamma z^t, & s = t \end{cases}$. The derivative $f_{\mu,c}(y(\gamma))' = \langle \nabla f_{\mu,c}(y(\gamma)), y(\gamma)' \rangle$ is hence zero except in the t-th place. Thus,

$$\begin{aligned} & f_{\mu,c}(x(\gamma))' \\ =\ & \langle \nabla_t f_{\mu,c}(y), -y^t + z^t \rangle \\ =\ & \langle [c \cdot y^t_\star(\gamma) + \mu^t - \nu_{A_t}\ 1], -y^t + z^t \rangle \\ =\ & \begin{array}{l} \langle [c \cdot y^t_\star + \mu^t - \nu_{A_t}\ 1], -y^t + z^t \rangle \\ +\gamma \langle c \cdot (-y^t_\star + z^t_\star), -y^t_\star + z^t_\star \rangle \end{array} \end{aligned} \tag{26}$$

Setting the above derivative zero yields

$$\gamma = \frac{\langle [c \cdot y^t_\star + \mu^t - \nu_{A_t}\ 1], y^t - z^t \rangle}{c \|y^t_\star - z^t_\star\|^2} .$$

Recalling that we require $\gamma \in [0, 1]$, we get the desired formula. ☐

B. Detailed experimental evaluation

In Table 2 we give the final lower bound obtained by each tested algorithm for every instance of every dataset we evaluated on. The averaged numbers are given in Table 1.

Table 2: Lower bound of each instance. † means method not applicable. **Bold** numbers indicate highest lower bound among competing methods.

Instance	FWMAP	CB	SA	MP
MRF				
protein folding				
1CKK	**-12840.23**	-12857.29	-12945.39	-12924.97
1CM1	**-12486.15**	-12499.21	-12591.23	-12488.10
1SY9	**-9193.38**	-9196.14	-9293.58	-9194.77
2BBN	**-12396.51**	-12461.89	-12585.85	-12417.20
2BCX	**-14043.57**	-14144.89	-14231.86	-14112.73
2BE6	**-13311.78**	-13381.35	-13410.24	-13438.23
2F3Y	**-14572.71**	-14619.70	-14672.71	-14641.60
2FOT	**-12049.52**	-12112.31	-12154.66	-12103.75
2HQW	**-13514.79**	-13573.99	-13610.14	-13539.69
2O60	**-13557.32**	-13664.00	-13718.71	-13565.42
3BXL	**-14125.86**	-14165.97	-14266.01	-14136.79
Discrete tomography				
2 projections				
0.10_0.10_2	**97.99**	97.94	96.46	†
0.20_0.20_2	**226.81**	226.66	222.05	†
0.30_0.30_2	**205.65**	205.25	194.49	†
0.40_0.40_2	**271.23**	270.99	253.94	†
0.50_0.48_87	**340.13**	339.98	315.41	†
0.60_0.58_28	**313.19**	312.80	288.73	†
0.70_0.67_47	**287.11**	286.83	246.04	†
0.80_0.76_72	**338.97**	338.78	290.73	†
0.90_0.85_63	**313.98**	313.77	246.63	†
4 projections				
0.10_0.10_2	**102.00**	101.55	99.50	†
0.20_0.20_2	**250.61**	250.02	245.30	†
0.30_0.30_2	**247.86**	246.44	233.65	†
0.40_0.40_2	**365.05**	364.00	346.89	†
0.50_0.48_87	**439.60**	435.50	412.32	†
0.60_0.58_28	**400.91**	400.05	368.05	†
0.70_0.67_47	**393.88**	392.57	371.80	†
0.80_0.76_72	**443.87**	440.91	413.42	†
0.90_0.85_63	**397.14**	395.93	358.60	†
6 projections				
0.10_0.10_2	**102.00**	**102.00**	101.82	†
0.20_0.20_2	**256.00**	255.85	254.74	†
0.30_0.30_2	**295.85**	292.28	272.74	†
0.40_0.40_2	**461.27**	456.70	433.89	†
0.50_0.48_87	**533.95**	526.86	494.29	†
0.60_0.58_28	**514.05**	507.34	474.61	†
0.70_0.67_47	**577.38**	566.15	530.47	†
0.80_0.76_72	**542.96**	534.01	488.62	†
0.90_0.85_63	**535.78**	518.67	468.60	†
sheep logan 64x64				
Logan_64_2	**582.52**	541.62	392.47	†
Logan_64_4	**871.58**	831.63	702.32	†

Table 2: Lower bound of each instance. † means method not applicable. **Bold** numbers indicate highest lower bound among competing methods.

Instance	FWMAP	CB	SA	MP
Logan_64_6	**1237.44**	1170.36	1011.00	†
sheep logan 256x256				
Logan_256_2	**3709.46**	3505.46	2599.41	†
Logan_256_4	**4888.25**	4739.40	976.29	†
Logan_256_6	**5142.48**	4832.85	-2463.81	†
Graph matching **6d scene flow**				
board	**-2262.66**	-2262.66	-2262.89	**-2262.66**
books	**-4179.79**	-4186.16	-4191.30	-4204.14
hammer	**-2125.87**	-2127.66	-2130.58	-2146.81
party	**-3648.03**	-3648.71	-3649.41	-3657.12
table	**-3340.59**	-3341.12	-3343.81	-3363.98
walking	**-1627.30**	-1627.34	-1627.58	-1627.79