Optimization for discrete graphical models

Paul Swoboda

February 18, 2019

1 Problem formulation

Definition 1 (Markov Random Field)**.** *Let $G = (V, E)$ be an undirected graph. Define the* label space *as $X = \prod_{i \in V} X_i$, where $|X_i| < \infty$ is a discrete set of labels. To each node $i \in V$ there exists a* unary potential *$\theta_i : X_i \to \mathbb{R}$ and to each edge $ij \in E$ there exists a* pairwise potential *$\theta_{ij} : X_i \times X_j \to \mathbb{R}$. We call the tuple (V, E, X, θ) a* Markov Random Field *or a* graphical model. *An element $x \in \mathcal{X}$ is called a* labeling. *We index its components by x_i.*

Definition 2 (MAP-inference)**.** *The problem $\min_{x \in X} \theta(x)$ with $\theta(x) = \sum_{i \in V} \theta_i(x_i) + \sum_{ij \in E} \theta_{ij}(x_i, x_j)$ is called the* Maximun-A-Posteriori (MAP)-inference *or energy minimization problem for the given MRF.*

2 Complexity of MAP-inference

Theorem 1. *The Hamiltonian cycle problem reduces to MAP-inference.*

Proof. Let the Hamiltonian cycle problem be given for graph $G' = (V', E')$. Define $G = (V, E)$ by $V = \{1, \ldots, |V'|\}$ and $E' = \binom{V'}{2}$, i.e. G is the full graph. Define $X_i = V'$ for all $i \in V$ and $\theta_i \equiv 0$ for all $i \in V$.

1. Let $\theta_{i,i+1}(s,t) = \begin{cases} 0, & (s,t) \in E' \\ \infty, & \text{otherwise} \end{cases}$ for $i = 1, \ldots, |V'|$, where we take $i+1 = 1$ for $i = |V'|$.

2. For all other $ij \in E$ let $\theta_{ij}(s,t) = \begin{cases} 0, & s \neq t \\ \infty, & \text{otherwise} \end{cases}$.

Then $x \in X$ with $\theta(x) < \infty$ defines a Hamiltonian cycle by construction: (x_i, x_{i+1}) corresponds to an edge due to in G (i) and no node is visited twice due to (ii). □

The above complexity result shows that probably no polynomial algorithm for MAP-inference exists. We will hence look at efficiently solvable subclasses (chains, submodular problems) and approximate algorithms for solving the general problem (LP-relaxation, message passing).

3 Chain models

For the rest of this section we will assume that $V = \{1, \dots, n\}$ and $E = \{\{i, i+1\}, i = 1, \dots, n-1\}$.

Definition 3. *Define*

$$F_j(s) = \min_{x \in X_1 \times \ldots \times X_{j-1} \times \{s\}} \sum_{i=1}^{j-1} \theta_i(x_i) + \sum_{i=1}^{j-1} \theta_{i,i+1}(x_i, x_{i+1})\,. \tag{1}$$

Clearly, $\min_{s \in X} F_n(s) + \theta_n(s) = \min_{x \in X} \theta(x)$. Also,

$$F_i(s) = \min_{t \in X_{i-1}} \left(F_{i-1}(t) + \theta_{i-1}(t) + \theta_{i-1,i}(t, s)\right)\,. \tag{2}$$

Algorithm 1: Dynamic programming for MAP-inference on chains

$F_1 \equiv 0$;
for $i = 2, \dots, n$ **do**
  for $s \in X_i$ **do**
    $F_i(s) = \min_{t \in X_{i-1}} \left(F_{i-1}(t) + \theta_{i-1}(t) + \theta_{i-1,i}(t, s)\right)$;
    Choose $r_i(s) \in \operatorname{argmin}_{t \in X_{i-1}} \left(F_{i-1}(t) + \theta_{i-1}(t) + \theta_{i-1,i}(t, s)\right)$;
  end
end
$E^* = \min_{s \in X_n} \left(F_n(s) + \theta_n(s)\right)$;
Choose $y_n \in \operatorname{argmin}_{s \in X_n} \left(F_n(s) + \theta_n(s)\right)$;
for $i = n, \dots, 2$ **do**
  Choose $y_{i-1} = r_i(y_i)$;
end

---

4 Linear programming

Definition 4 (Polyhedron)**.** *Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ be given. We call the set $\{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$ the polyhedron associated to (A, b).*

Definition 5 (Standard simplex)**.** *We call the polyhedron*

$$\mathit{\Delta}_n = \{x \in \mathbb{R}^n_+ : \sum_{i=1}^{n} x_i = 1\} \tag{3}$$

the n-dimensional standard simplex, or just simplex for conciseness.

Definition 6 (Convex hull)**.** *Let $\delta^i \in \mathbb{R}^n$, $i = 1, \dots, N$ be points in $\mathbb{R}^n$. Define the convex hull as*

$$\operatorname{conv}\{\delta^1, \dots, \delta^N\} = \{x : \exists p \in \delta_N \text{ s.t. } x = \sum_{i=1}^{N} p_i \cdot \delta^i\}\,. \tag{4}$$

Lemma 1. *Let $a \in \mathbb{R}^n$. Then*

$$\min_{i=1,\dots,n}\{a_i\} = \min_{p\in\Delta_n}\sum_{i=1}^{n} p_i a_i = \min_{\mu\in\operatorname{conv}\{a_1,\dots,a_n\}} \mu \tag{5}$$

4.1 Marginal polytope

From now on we assume that the label space for all nodes $i \in V$ is $X_i = \{1, \dots, |X_i|\}$.

Definition 7 (Labeling mapping)**.** 1. *Define the node mapping for all $i \in V$ and all $x_i \in X_i$*

$$\delta_i(x_i) = (0, \dots, 0, \underbrace{1}_{\text{i-th position}}, 0, \dots, 0\}^\top . \tag{6}$$

2. *Define the edge mapping for all $ij \in E$ and all $x_i \in X_i$, $x_j \in X_j$ by*

$$\delta_{ij}(x_i, x_j) = vec(\begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ \vdots & & 1 & & \vdots \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}), \tag{7}$$

with a 1 in the x_i-th column and x_j-th row and 0 everywhere else in the matrix.

3. *Define the overall labeling mapping $\delta : X \to \{0,1\}^I$, where $I = \sum_{i\in V}|X_i| + \sum_{ij\in}|X_i| \cdot |X_j|$ by concatenating all node and edge mappings.*

4. *Similarly, define the cost vector by (abusing notation) $\theta_i(x_i) = \langle\theta_i, \delta_i(x_i)\rangle$ for all $i \in V$, $x_i \in X_i$ and $\theta_{ij}(x_i, x_j) = \langle\theta_{ij}, \delta_{ij}(x_i, x_j)\rangle$ for all $ij \in E$, $x_i \in X_i$, $x_j \in X_j$. and denote again the concatenated vector by $\theta \in \mathbb{R}^I$.*

Definition 8 (Marginal Polytope)**.** *Define $\mathcal{M} = \operatorname{conv}\{\delta(x) : x \in X\}$.*

Proposition 1.

$$\min_{x\in X}\theta(x) = \min_{\mu\in\mathcal{M}}\langle\theta, \mu\rangle . \tag{8}$$

Proof. "$\leq$": Let $\mu \in \mathcal{M}$. Then $\mu = \sum_{x\in X} p_x\delta(x)$ and hence $\langle\theta, \mu\rangle = \sum_{x\in X} p_x\langle\theta, \delta(x)\rangle \geq \min_{x\in X}\langle\theta(x), \delta(x)\rangle = \min_{x\in X}\theta(x)$.

"$\geq$": Choose $\mu = \delta(x^*)$ for some $x^* \in \operatorname{argmin}_{x\in X}\theta(x)$. □

4.2 Local marginal polytope

Since MAP-inference is NP-hard, the marginal polytope has no known characterization as a polytope with either polynomially many inequalities describing it nor does there exist a known polynomial time separation routine to determine whether $x \in \mathcal{M}$. Hence, we will study a simpler polytope $\mathcal{L}$ with $\mathcal{M} \subset \mathcal{L}$ that has a simple characterization but that still allows to obtain good results in practice.

Definition 9 (Local marginal polytope)**.** *Define*

$$\mathcal{L} = \left\{ \mu \in \mathbb{R}^I : \begin{array}{cc} \sum_{x_i \in X_i} \mu_i(x_i) = 1, & \forall i \in V \\ \sum_{x_i \in X_i, x_j \in X_j} \mu_{ij}(x_i, x_j) = 1, & \forall ij \in E \\ \sum_{x_i \in X_i} \mu_{ij}(x_i, x_j) = \mu_i(x_i), & \forall ij \in E, x_i \in X_i \\ \sum_{x_j \in X_j} \mu_{ij}(x_i, x_j) = \mu_j(x_j), & \forall ij \in E, x_j \in X_j \end{array} \right\} . \tag{9}$$

Proposition 2.

$$\mathcal{M} \subseteq \mathcal{L} . \tag{10}$$

Corollary 1.

$$\min_{\mu \in \mathcal{L}} \langle \theta, \mu \rangle \leq \min_{\mu \in \mathcal{M}} \langle \theta, \mu \rangle = \min_{\mu \in \mathcal{L} \cap \{0,1\}^I} \langle \theta, \mu \rangle . \tag{11}$$

4.3 Linear programming duality

Let (A, b) describe a polyhedron. For some objective vector c we call $\min_{\{x \geq 0 : Ax = b\}} \langle c, x \rangle$ the *primal problem* and $\max_{\{y : A^\top y \leq c\}} \langle b, y \rangle$ the *dual problem*.

Proposition 3 (Weak duality)**.**

$$\min_{\{x \geq 0 : Ax = b\}} \langle c, x \rangle \geq \max_{y : A^\top y \leq c} \langle b, y \rangle . \tag{12}$$

Proof. Let x and y be feasible to the primal resp. dual problem. Then

$$\begin{aligned} & \langle c, x \rangle \geq \langle b, y \rangle \\ \Leftrightarrow & \langle c, x \rangle - \langle b, y \rangle \geq 0 \\ \Leftrightarrow & \langle c, x \rangle - \langle A^\top x, y \rangle \geq 0 \\ \Leftrightarrow & \langle \underbrace{x}_{\geq 0}, \underbrace{c - A^\top y}_{\geq 0} \rangle \geq 0 . \end{aligned} \tag{13}$$

□

Proposition 4 (Strong duality)**.**

$$\min_{\{x \geq 0 : Ax = b\}} \langle c, x \rangle = \max_{y : A^\top y \leq c} \langle b, y \rangle . \tag{14}$$

4.4 Linear programming duality for the local marginal polytope

We apply linear programming duality to optimizing over the linear polytope relaxation $\mathcal{L}$.

$$\begin{array}{lr} \min_\mu \langle \theta, \mu \rangle & \max_{z,\phi} \sum_{i \in V} z_i + \sum_{ij \in E} z_{ij} \\ \sum_{x_i \in X_i} \mu_i(x_i) = 1 & z_i \\ \sum_{x_i \in X_i, x_j \in X_j} \mu_{ij}(x_i, x_j) = 1 & z_{ij} \\ \sum_{x_j \in X_j} \mu_{ij}(x_i, x_j) - \mu_i(x_i) = 0 & \phi_{i \to j}(x_i) \\ \sum_{x_i \in X_i} \mu_{ij}(x_i, x_j) - \mu_j(x_j) = 0 & \phi_{j \to i}(x_j) \\ \mu_i(x_i) \geq 0 & z_i - \sum_{j : ij \in E} \phi_{i \to j}(x_i) \leq \theta_i(x_i) \\ \mu_{ij}(x_i, x_j) \geq 0 & z_{ij} + \phi_{i \to j}(x_i) + \phi_{j \to i}(x_j) \leq \theta_{ij}(x_i, x_j) \end{array} \tag{15}$$

Definition 10 (Reparametrization)**.** *For any dual variables ϕ we call*

$$\theta_i^{\phi}(x_i) = \theta_i(x_i) + \sum_{j:ij\in E} \phi_{i\to j}(x_j) \tag{16}$$

and

$$\theta_{ij}^{\phi}(x_i, x_j) = \theta_{ij}(x_i, x_j) - \phi_{i\to j}(x_i) - \phi_{j\to i}(x_j) \tag{17}$$

reparametrized unary resp. dual potentials.

We can succinctly rewrite the dual optimization problem over the local polytope relaxation in terms of reparametrizations as

$$\max_{z,\phi} \sum_{i\in V} z_i \text{ s.t. } \begin{array}{ll} z_i \leq \theta_i^{\phi}(x_i), & \forall i \in V, x_i \in X_i \\ z_{ij} \leq \theta_{ij}^{\phi}(x_i, x_j), & \forall ij \in E, x_i \in X_i, x_j \in X_j \end{array} \tag{18}$$

and even shorter as

$$\max_{\phi} \sum_{i\in V} \min_{x_i\in X_i} \left\{\theta_i^{\phi}(x_i)\right\} + \sum_{ij\in E} \min_{x_i\in X_i, x_j\in X_j} \left\{\theta_{ij}^{\phi}(x_i, x_j)\right\} \tag{19}$$

4.5 Optimality conditions

Let $\mu^* \in \operatorname{argmin}_{\mu\in\mathcal{L}} \langle\theta, \mu\rangle$ be a primal optimal solution for the local polytope relaxation. Let

$$\begin{array}{rl} (\phi^*, z^*) \in \underset{\phi,z}{\operatorname{argmax}} & \sum_{i\in V} z_i + \sum_{ij\in E} z_{ij} \\ \text{s.t.} & z_i \leq \theta_i^{\phi}(x_i) \quad \forall i \in V, x_i \in X_i \\ & z_{ij} \leq \theta_{ij}^{\phi}(x_i, x_j) \quad \forall ij \in E, x_i \in X_i, x_j \in X_j \end{array} \tag{20}$$

be a dual optimal solution. Write the constraint matrix defining the local polytope relaxation as $\mathcal{L} = \{\mu : A\mu = \begin{pmatrix} \mathbb{1} \\ 0 \end{pmatrix}$. Then the primal/dual LP-optimality conditions applied to the local polytope read

$$\begin{aligned} 0 =& \langle\theta, \mu^* > -\langle \begin{pmatrix} \mathbb{1} \\ 0 \end{pmatrix}, \begin{pmatrix} z^* \\ \phi^* \end{pmatrix}\rangle \\ =& \langle\theta, \mu^*\rangle - \langle A\mu^*, \begin{pmatrix} z^* \\ \phi^* \end{pmatrix}\rangle \\ =& \langle\mu^*, \theta - A^\top \begin{pmatrix} z^* \\ \phi^* \end{pmatrix}\rangle \\ =& \sum_{i\in V} \langle\mu_i, \underbrace{\theta_i^{\phi^*} - z_i^*}_{\geq 0}\rangle + \sum_{ij\in E} \langle\mu_{ij}^*, \underbrace{\theta_{ij}^{\phi^*} - x_{ij}^*}_{\geq 0}\rangle \end{aligned} \tag{21}$$

This implies that

$$\mu_i^*(x_i) > 0 \Rightarrow \theta_i^{\phi^*}(x_i) = \min_{x_i'\in X_i} \{\theta_i^{\phi^*}(x_i')\} = z_i \tag{22}$$

and

$$\mu_{ij}^*(x_i, x_j) > 0 \Rightarrow \theta_{ij}^{\phi^*}(x_i, x_j) = \min_{x_i'\in X_i, x_j'\in X_j} \{\theta_{ij}^{\phi^*}(x_i', x_j')\} = z_{ij} \tag{23}$$

5 Dual block coordinate ascent

Remark 1. *The class of algorithms we will present are an instance of the family of dual block coordinate ascent techniques. In the literature, the algorithms are also referred to by message passing and belief propagation.*

Definition 11 (Elementary steps)**.** *Define the elementary message computation from node $i \in V$ to edge $ij \in E$ as*

$$\begin{aligned} \operatorname*{msg}_{i\to ij} : \mathbb{R}^{|X_i|} &\to \mathbb{R}^{|X_i|} \\ \theta_i^{\phi} &\mapsto \theta_i^{\phi}\,. \end{aligned} \tag{24}$$

In other words, $\mathrm{msg}_{i\to ij}$ *is the identity.*

Define the elementary message computation from edge $ij \in E$ to node $i \in V$ as

$$\begin{aligned} \operatorname*{msg}_{ij\to i} : \mathbb{R}^{|X_i|\times|X_j|} &\to \mathbb{R}^{|X_i|} \\ \theta_{ij}^{\phi} &\mapsto \left(\min_{x_j\in X_j}\{\theta_{ij}(x_i,x_j)\}\right)_{x_i\in X_i} \end{aligned} \tag{25}$$

and from edge $ij \in E$ to node $j \in V$ as

$$\begin{aligned} \operatorname*{msg}_{ij\to j} : \mathbb{R}^{|X_i|\times|X_j|} &\to \mathbb{R}^{|X_j|} \\ \theta_{ij}^{\phi} &\mapsto \left(\min_{x_i\in X_i}\{\theta_{ij}(x_i,x_j)\}\right)_{x_j\in X_j} \end{aligned} \tag{26}$$

We define a basic MPLP step in Algorithm 2.

Algorithm 2: MPLP(ij)

I {
$\Delta_i \leftarrow \mathrm{msg}_{i\to ij}(\theta_i^{\phi})$;
$\phi_{i\to j} \leftarrow \phi_{i\to j} - \Delta_i$;

II {
$\Delta_j \leftarrow \mathrm{msg}_{j\to ij}(\theta_j^{\phi})$;
$\phi_{j\to i} \leftarrow \phi_{j\to i} - \Delta_j$;

III {
$\Delta_i' \leftarrow \mathrm{msg}_{ij\to i}(\theta_{ij}^{\phi})$;
$\Delta_j' \leftarrow \mathrm{msg}_{ij\to j}(\theta_{ij}^{\phi})$;
$\phi_{i\to j} \leftarrow \phi_{i\to j} + \frac{1}{2}\Delta_i'$;
$\phi_{j\to i} \leftarrow \phi_{j\to i} + \frac{1}{2}\Delta_j'$;

Proposition 5. *MPLP(ij) improves the dual lower bound for every $ij \in E$.*

Proof. We will prove that steps I, II and III in Algorithm 2 all individually improve the dual lower bound. Since only messages $\phi_{i\to j}$ and $\phi_{j\to i}$ are affected, it is enough to consider

$$\min_{x_i\in X_i}\{\theta_i(x_i)\} + \min_{x_j\in X_j}\{\theta_j(x_j)\} + \min_{x_{ij}\in X_{ij}}\{\theta_{ij}(x_{ij})\} \tag{27}$$

before and after operations I, II and III. Let ϕ° be the dual variables before and ϕ the dual variable after the respective steps I, II and III.

I: We have $\min_{x_i \in X_i} \left\{ \theta_i^{\phi} \right\} \equiv 0$.

$$
\begin{aligned}
&\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi}(x_i, x_j) \right\} \\
=&\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) + \theta_i^{\phi^\circ}(x_i) \right\} \\
\geq&\underbrace{\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) \right\} + \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) \right\}}_{\text{lower bound brefore I}}
\end{aligned}
\tag{28}
$$

II: Analoguous to I.

III:

$$
\begin{aligned}
&\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta^{\phi}(x_i, x_j) \right\} \\
=&\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta^{\phi^\circ}(x_i, x_j) - \frac{1}{2}\Delta_i'(x_i) - \frac{1}{2}\Delta_j'(x_j) \right\} \\
\geq&\min_{x_i \in X_i, x_j \in X_j} \left\{ \frac{1}{2}\theta^{\phi^\circ}(x_i, x_j) - \frac{1}{2}\Delta_i'(x_i) \right\} \\
&+ \min_{x_i \in X_i, x_j \in X_j} \left\{ \frac{1}{2}\theta^{\phi^\circ}(x_i, x_j) - \frac{1}{2}\Delta_j'(x_j) \right\} \\
=&0\,.
\end{aligned}
\tag{29}
$$

Also,

$$
\begin{aligned}
&\min_{x_i \in X_i} \left\{ \theta_i^{\phi}(x_i) \right\} \\
&= \min_{x_i \in X_i} \left\{ \frac{1}{2}\Delta_i' \right\} \\
&= \min_{x_i \in X_i} \left\{ \frac{1}{2} \min_{x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) \right\} \right\} \\
&= \min_{x_i \in X_i, x_j \in X_j} \left\{ \frac{1}{2}\theta_{ij}^{\phi^\circ}(x_i, x_j) \right\}
\end{aligned}
\tag{30}
$$

and

$$
\begin{aligned}
&\min_{x_i \in X_i} \left\{ \theta_i^{\phi}(x_i) \right\} \\
&= \min_{x_i \in X_i} \left\{ \frac{1}{2}\Delta_i' \right\} \\
&= \min_{x_i \in X_i} \left\{ \frac{1}{2} \min_{x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) \right\} \right\} \\
&= \min_{x_i \in X_i, x_j \in X_j} \left\{ \frac{1}{2}\theta_{ij}^{\phi^\circ}(x_i, x_j) \right\}
\end{aligned}
\tag{31}
$$

□

The overall MPLP Algorithm 3 works by iterating over all edges $ij \in E$ and performing the basic MPLP step Algorithm 2.

Algorithm 3: MPLP

```
for t = 1, ... do
    for ij ∈ E do
        MPLP(ij);
    end
end
```

There is a family of algorithms that work similar to MPLP. The order of message updates is however reversed: First come pairwise to unary messagesa and afterwards come unary to pairwise messages. We call a basic step of such an algorithm a diffusion step as detailed in Algorithm 4. For easy exposition define the neighborhood of any node $i \in V$ as

$$N_i = \{j \in V : ij \in E\} . \tag{32}$$

Algorithm 4: Diffusion step

<u>DS</u> (i, R_i, ω);

Input : $i \in V,\ R_i \subset N_i,\ \omega \in \mathbb{R}_+^{N_i},\ \sum_{j \in N_i} \omega_j \leq 1$.

I:

for $j \in R_i$ **do**

  $\Delta_j \leftarrow \mathrm{msg}_{ij \to i}(\theta_{ij}^{\phi})$;

  $\phi_{i \to j} \leftarrow \phi_{i \to j} + \Delta_j$;

end

II:

for $j \in N_i$ **do**

  $\Delta'_j \leftarrow \mathrm{msg}_{i \to ij}(\theta_i^{\phi})$;

end

for $j \in N_i$ **do**

  $\phi_{i \to j} \leftarrow \phi_{i \to j} - \omega_j \Delta'_j$;

end

Proposition 6. *For any $i \in V$, $R_i \subset N_i$ and $\omega \in \mathbb{R}_+^{N_i} : \sum_{j \in N_i} \omega_j \leq 1$ the basic diffusion step $DS(i, R_i, \omega)$ is monotonous.*

Proof. Similarly as in the proof of Proposition 5 we denote by ϕ° and ϕ the dual variables before and after step I resp. II. Also we prove the property individually for step I and II.

I: We have for all $j \in R_i$

$$\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi} \right\} \equiv 0\,. \tag{33}$$

Also

$$\begin{aligned}
&\min_{x_i \in X_i} \left\{ \theta_i^{\phi}(x_i) \right\} \\
&\qquad = \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) + \Delta_i(x_i) \right\} \\
&\qquad \geq \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) \right\} + \min_{x_i \in X_i} \left\{ \min_{x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) \right\} \right\} \\
&\qquad \geq \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) \right\} + \min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) \right\}
\end{aligned} \tag{34}$$

II:

$$\begin{aligned}
&\min_{x_i \in X_i} \left\{ \theta_i^{\phi}(x_i) \right\} \\
&= \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) - \sum_{j \in N_i} \Delta'_j(x_i) \right\} \\
&= \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) - \sum_{j \in N_i} \theta_i^{\phi^\circ}(x_i) \right\} \\
&= (1 - \sum_{j \in N_i} \omega_j) \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) \right\}
\end{aligned} \tag{35}$$

For any $j \in N_i$ we have

$$\begin{aligned}
&\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi}(x_i, x_j) \right\} \\
= &\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) + \omega_j \Delta'_j(x_i) \right\} \\
\geq &\min_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi^\circ}(x_i, x_j) \right\} + \omega_j \min_{x_i \in X_i} \left\{ \theta_i^{\phi^\circ}(x_i) \right\}
\end{aligned} \tag{36}$$

□

Example 1. *Thera are multiple choices for the sets R_i and ω for the basic diffusion step $DS(i, R_i, \omega)$.*

(i) Min-sum diffusion: Choose for any $i \in V$ the set $R_i = N_i$ and $\omega = \frac{1}{|N_i|}\mathbb{1}$.

(ii) Tree-reweighted message passing: Choose some order on nodes $V = \{1, \ldots, |V|\}$. We visit the nodes in the given order, see Algorithm 5.

After computing a forward pass, we invert the order on V and perform the above procedure.

Algorithm 5: Forward pass of TRWS

for $i = 1, \ldots, n$ **do**

  $R_i = \{j \in N_i : j < i\}$;

  $\omega_j = \begin{cases} 0, & j \in R_i \\ \frac{1}{\max(|R_i|, |N_i \backslash R_i|}, j \notin R_i \end{cases}$;

  $DS(i, R_i, \omega)$

end

(iii) TRWS primal rounding: After step I in Algorithm 4 and before step II we do the following:

$$x_i^* \in \operatorname*{argmin}_{x_i \in X_i} \left\{ \theta_i^\phi(x_i) + \sum_{k \in N_i : k < i} \theta_{ki}(x_k^*, x_i) \right\} . \tag{37}$$

6 Submodularity

We assume throughout this section that $X_i = \{0, 1\}$ for all $i \in V$.

Definition 12 (Submodularity)**.** *A pairwise potential $\theta_{ij} : \{0,1\}^2 \to \mathbb{R}$ is submodular iff*

$$\theta_{ij}(0,0) + \theta_{ij}(1,1) \leq \theta_{ij}(0,1) + \theta_{ij}(1,0) . \tag{38}$$

An energy $\sum_{i \in V} \theta_i(x_i) + \sum_{ij \in E} \theta_{ij}(x_i, x_j)$ is submodular if all its pairwise potentials are.

Example 2 (Ising model)**.** *$\theta_{ij}(x_i, x_j) = \alpha \mathbb{1}_{x_i != x_j}$ is submodular if $\alpha \geq 0$.*

Lemma 2. *Define the minimum operation $x, y \mapsto x \wedge y = \min(x, y)$ and the maximum operation as $x, y \mapsto x \vee y = \max(x, y)$. Then for any $x_i, y_i \in X_i$ and $x_j, y_j \in X_j$ and any submodular potential $\theta_i j$ it holds that*

$$\theta_{ij}(x_i \wedge y_i, x_j \wedge y_j) + \theta_{ij}(x_i \vee y_i, x_j \vee y_j) \leq \theta_{ij}(x_i, x_j) + \theta_{ij}(y_i, y_j) . \tag{39}$$

Lemma 3. *Let θ be an MRF energy and ϕ any dual variables. If θ is submodular, so is θ^ϕ.*

Proof. For any $ij \in E$ we have

$$\begin{aligned} & \theta_{ij}^\phi(0,0) + \theta_{ij}^\phi(1,1) \\ =& \theta_{ij}(0,0) - \phi_{i \to j}(0) - \phi_{j \to i}(0) + \theta_{ij}(1,1) - \phi_{i \to j}(1) - \phi_{j \to i}(1) \\ \leq& \underbrace{\theta_{ij}(0,1) - \phi_{i \to j}(0) - \phi_{j \to i}(1)}_{= \theta_{ij}^\phi(0,1)} + \underbrace{\theta_{ij}(1,0) - \phi_{i \to j}(1) - \phi_{j \to i}(0)}_{= \theta_{ij}^\phi(1,0)} \end{aligned} \tag{40}$$

□

When having solved the primal and dual problem over the local polytope relaxation, we still need to reconstruct a primal solution. We will show that it can be done such that solution will fulfill the primal/dual optimality conditions. Hence, the reconstructed primal solution is optimal.

Definition 13. *Let $\mu^* \in \operatorname{argmin}_{\mu \in \mathcal{L}} \langle \theta, \mu \rangle$ be an optimal primal solution of the local polytope relaxation and let ϕ^* be an optimal dual solution. Define for every $i \in V$ the space of locally optimal labels for the unary potentials as*

$$S_i = \operatorname*{argmin}_{x_i \in X_i} \left\{ \theta_i^{\phi^*}(x_i) \right\} . \tag{41}$$

Denote for every edge $ij \in E$ the space of locally optimal labels for the pairwise potentials as

$$S_{ij} = \operatorname*{argmin}_{x_i \in X_i, x_j \in X_j} \left\{ \theta_{ij}^{\phi^*}(x_i, x_j) \right\} . \tag{42}$$

For every node $i \in V$ denote the support setf μ_i^* as

$$O_i = \{ x_i \in X_i : \mu_i^*(x_i) > 0 \} . \tag{43}$$

Proposition 7. *For all $i \in V$ choose x_i^* as the greatest label from O_i. Then x^* is optimal.*

Proof. We check primal/dual optimality conditions, i.e. that $x_i^* \in S_i$ and $(x_i^*, x_j^*) \in S_{ij}$.

- For $i \in V$ we have by definition of O_i and the primal/dual optimality conditions that $x_i \in S_i$.
- For $ij \in E$ we have that for i there exists a $x_j' \in O_j$ with $(x_i^*, x_j') \in S_{ij}$, since it holds that $\mu_i^*(x_i^*) > 0$ and therefore there must exist a $x_j' \in X_j$ with $\mu_{ij}^*(x_i^*, x_j') > 0$ (due to the marginalization constraints of the local polytope relaxation). $\mu_{ij}^*(x_i^*, x_j') > 0$ (due to the marginalization constraints of the local polytope relaxation). Further $\mu^*(x_i^*, x_j') > 0$ implies that $(x_i^*, x_j') \in S_{ij}$ due to primal/dual optimality conditions. Last, $x_j' \in O_j$ again due to the marginalization constraints of the local polytope relaxation. Similarly, there exists $x_i' \in O_i$ such that $(x_i', x_j^*) \in S_{ij}$. Then

  $$\theta^{\phi^*}(x_i', x_j') \geq \theta_{ij}^{\phi^*}(x_i^*, x_j') = \theta_{ij}^{\phi^*}(x_i', x_j^*) \leq \theta_{ij}^{\phi^*}(x_i^*, x_j^*) . \tag{44}$$

  Since $x_i' \in O_i$ we have $x_i' \leq x_i^*$ and likewise $x_j' \leq x_j^*$ due to the choice of x^*. Hence

  $$\begin{aligned} &\theta_{ij}^{\phi^*}(x_i^*, x_j^*) + \theta_{ij}^{\phi^*}(x_i', x_j') \\ &\quad = \theta_{ij}^{\phi^*}(x_i^* \wedge x_i', x_j^* \wedge x_j') + \theta_{ij}^{\phi^*}(x_i^* \vee x_i', x_j^* \vee x_j') \\ &\quad\quad \leq \theta_{ij}^{\phi^*}(x_i^*, x_j') + \theta_{ij}^{\phi^*}(x_i', x_j^*) . \end{aligned} \tag{45}$$

□