Appendix: Worst-case Optimal Submodular Extensions for Marginal Estimation

1 Proofs for Potts Model Extension

Remark 1 We show using induction over the number of variables that with 1-of-L encoding for Potts,

$$\sum_{A \in \mathcal{M}} \exp(-s(A)) = \prod_{a=1}^{N} \sum_{i=1}^{L} \exp(-s_{ai}). \tag{1}$$

Proof. Let t be the number of variables, V^t be the corresponding ground set and $\mathcal{M}^t$ be the sets corresponding to valid labelings. Equation (1) clearly holds for $t = 1$.

Let us assume that the relation holds for $t = N$, that is,

$$\sum_{A^N \in \mathcal{M}^N} \exp(-s(A^N)) = \prod_{a=1}^{N} \sum_{i=1}^{L} \exp(-s_{ai}) \tag{2}$$

For $t = N + 1$,

$$\begin{aligned}
\sum_{A^{N+1} \in \mathcal{M}^{N+1}} \exp(-s(A^{N+1})) &= \sum_{i=1}^{L} \sum_{A^N \in \mathcal{M}^N} \exp(-s(A^N) - s_{N+1,i}) \\
&= \sum_{i=1}^{L} \exp(-s_{N+1,i}) \sum_{A^N \in \mathcal{M}^N} \exp(-s(A^N)) \\
&= \sum_{i=1}^{L} \exp(-s_{N+1,i}) \prod_{a=1}^{N} \sum_{i=1}^{L} \exp(-s_{ai}) \\
&= \prod_{a=1}^{N+1} \sum_{i=1}^{L} \exp(-s_{ai})
\end{aligned} \tag{3}$$

□

Remark 2 Given any submodular extension $F(.)$ of a Potts energy function $E(.)$, its Lovasz extension $f(.)$ defines an LP relaxation of the MAP problem for $E(.)$ as

$$\min_{\mathbf{y} \in \Delta} f(\mathbf{y}). \tag{4}$$

Proof. By definition of a submodular extension and the Lovasz extension, $E(\mathbf{x}) = F(A_{\mathbf{x}}) = f(1_{A_{\mathbf{x}}})$ for all valid labelings $\mathbf{x}$. Also, from property 1, $f(\mathbf{y})$ is maximum of linear functions. Hence, $f(\mathbf{y})$ is a piecewise linear relaxation of $E(\mathbf{x})$.

The domain Δ is a polytope formed by union of N probability simplices

$$\Delta = \{\mathbf{y}_a \in \mathbb{R}^L | \mathbf{y}_a \succeq 0 \text{ and } \langle \mathbf{1}, \mathbf{y}_a \rangle = 1\} \tag{5}$$

With objective as maximum of linear functions and domain as a polytope, we have an LP relaxation of the corresponding MAP problem. □

Proposition 1. *In the limit $T \to 0$, the following problem for Potts energies*

$$\min_{\mathbf{s} \in EP(F)} \quad g_T(\mathbf{s}) = \sum_{a=1}^{N} T \cdot \log \sum_{i=1}^{L} \exp(-\frac{s_{ai}}{T}). \tag{6}$$

becomes

$$-\min_{\mathbf{y} \in \Delta} f(\mathbf{y}). \tag{7}$$

Proof. In the limit of $T \to 0$, we can rewrite the above problem as

$$\min_{\mathbf{s} \in EP(F)} \quad \sum_{a=1}^{N} \max_i(-s_{ai}). \tag{8}$$

In vector form, the problem becomes

$$\min_{\mathbf{s} \in EP(F)} \max_{\mathbf{y} \in \Delta} -\langle \mathbf{y}, \mathbf{s} \rangle \tag{9}$$

$$= - \max_{\mathbf{s} \in EP(F)} \min_{\mathbf{y} \in \Delta} \langle \mathbf{y}, \mathbf{s} \rangle \tag{10}$$

Δ is the union of N probability simplices:

$$\Delta = \{\mathbf{y}_a \in \mathbb{R}^L | \mathbf{y}_a \succeq 0 \text{ and } \langle \mathbf{1}, \mathbf{y}_a \rangle = 1\} \tag{11}$$

where $\mathbf{y}_a$ is the component of $\mathbf{y}$ corresponding to the a-th variable. By the minimax theorem for LP, we can reorder the terms:

$$-\min_{\mathbf{y} \in \Delta} \max_{\mathbf{s} \in EP(F)} \langle \mathbf{y}, \mathbf{s} \rangle \tag{12}$$

Recall that $\max_{\mathbf{s} \in EP(F)} \langle \mathbf{y}, \mathbf{s} \rangle$ is the value of the Lovasz extension of F at $\mathbf{y}$, that is, $f(\mathbf{y})$. Hence, as $T \to 0$, the marginal inference problem converts to minimising the Lovasz extension under the simplices constraint:

$$-\min_{\mathbf{y} \in \Delta} f(\mathbf{y}) \tag{13}$$

□

Proposition 2. *The objective function $E(\mathbf{y})$ of the LP relaxation (P-LP) is the Lovasz extension of $F_{Potts}(A) = \sum_{i=1}^{L} F_i(A)$, where*

$$F_i(A) = \sum_a \phi_a(i)[|A \cap \{v_{ai}\}| = 1] + \sum_{(a,b)\in\mathcal{N}} \frac{w_{ab}}{2} \cdot [|A \cap \{v_{ai}, v_{bi}\}| = 1]. \tag{14}$$

Proof. Since F_{Potts} is sum of Ising models F_i, we first focus on a particular label i and then generalize. Consider a graph with only two variables X_a and X_b with an edge between them. The ground set in this case is $\{v_{ai}, v_{bi}\}$. Let the corresponding relaxed indicator variables be $\mathbf{y} = \{y_{aj}, y_{bj}\}$, such that $y_{ai}, y_{bi} \in [0,1]$ and assume $y_{ai} > y_{bi}$. The Lovasz extension is:

$$\begin{aligned} f(\mathbf{y}) &= y_{ai} \cdot [F_i(\{v_{ai}\}) - F_i(\{\})] + y_{bi} \cdot [F_i(\{v_{ai}, v_{bi}\}) - F_i(\{v_{ai}\})] \\ &= y_{ai} \cdot [\Big(\phi_a(j) + \frac{w_{ab}}{2}\Big) - 0] + y_{bi} \cdot [(\phi_a(j) + \phi_b(j)) - \Big(\phi_a(j) + \frac{w_{ab}}{2}\Big)] \\ &= \phi_a(j) \cdot y_{ai} + \phi_b(j) \cdot y_{bi} + \frac{w_{ab}}{2} \cdot (y_{ai} - y_{bi}) \end{aligned} \tag{15}$$

In general for both orderings of y_{ab} and y_{bi}, we can write

$$f(\mathbf{y}) = \phi_a(j) \cdot y_{ai} + \phi_b(j) \cdot y_{bi} + \frac{w_{ab}}{2} \cdot |y_{ai} - y_{bi}| \tag{16}$$

Extending Lovasz extension (equation (16)) to all variables and labels gives $E(\mathbf{y})$ in (P-LP). □

2 Proofs for Hierarchical Potts Model Extension

Transformed Tightest LP Relaxation We take (T-LP) and rewrite it using indicator variables for all labels and meta-labels. Let $\mathcal{R}$ denote the set of all labels and meta-labels, that is, all nodes in the tree apart from the root. Also, let $\mathcal{L}$ denote the set of labels, that is, the leaves of the tree. Let T_i denote the subtree which is rooted at the i-th node. We introduce an indicator variable $z_{ai} \in \{0,1\}$, where

$$z_{ai} = \begin{cases} y_{ai} & \text{if } i \in \mathcal{L} \\ y_a(T_i) & \text{if } i \in \mathcal{R} - \mathcal{L} \end{cases} \tag{17}$$

We need to extend the definition of unary potentials to the expanded label space as follows:

$$\text{where } \phi'_a(i) = \begin{cases} \phi_a(i) & \text{if } i \in \mathcal{L} \\ 0 & \text{if } i \in \mathcal{R} - L \end{cases} \tag{18}$$

We can now rewrite problem (T-LP) in terms of new indicator variables z_{ai}:

$$\begin{aligned} \text{(T-LP-FULL)} \quad \min \widetilde{E}(\mathbf{z}) = \sum_{i\in\mathcal{R}} \sum_{a\in\mathcal{X}} \phi'_a(i) \cdot z_{ai} + \\ \sum_{i\in\mathcal{R}} \sum_{(a,b)\in\mathcal{N}} w_{ab} \cdot l_{T_i} \cdot |z_{ai} - z_{bi}| \\ \text{such that} \quad \mathbf{z} \in \Delta' \end{aligned} \tag{19}$$

where Δ' is the convex hull of the vectors satisfying

$$\sum_{i \in \mathcal{L}} z_{ai} = 1, \;\; z_{ai} \in \{0,1\} \;\; \forall a \in \mathcal{X}, i \in \mathcal{L} \tag{20}$$

$$\text{and } \; z_{ai} = \sum_{j \in L(T_i)} z_{aj}. \;\; \forall a \in \mathcal{X}, i \in \mathcal{R} - \mathcal{L} \tag{21}$$

Constraint (21) ensures consistency among labels and meta-labels, that is, if a label is assigned then all the meta-labels which lie on the path from the root to the label should be assigned as well. We are now going to identify a suitable set encoding and the worst-case optimal submodular extension using (T-LP-FULL).

Remark 3 Given any submodular extension $F(.)$ of a hierachical Potts energy function $E(.)$, its Lovasz extension defines an LP relaxation of the corresponding MAP estimation problem as

$$\min_{\mathbf{z} \in \Delta'} f(\mathbf{z}). \tag{22}$$

Proof. By definition of a submodular extension and the Lovasz extension, $E(\mathbf{x}) = F(A_{\mathbf{x}}) = f(1_{A_{\mathbf{x}}})$ for all valid labelings $\mathbf{x}$. Also, from property 1, $f(\mathbf{y})$ is maximum of linear functions. Hence, $f(\mathbf{y})$ is a piecewise linear relaxation of $E(\mathbf{x})$.

We can write the domain Δ' as

$$\Delta' = \{\mathbf{y}_a \in \mathbb{R}^M | \mathbf{y}_a \succeq 0, \;\; \langle \mathbf{1}, \mathbf{y}_a^{label} \rangle = 1, \;\; \mathbf{y}_a(p_{ai}) = \mathbf{1} \text{ or } \mathbf{y}_a(p_{ai}) = \mathbf{0} \forall i \in [1, L]\} \tag{23}$$

where $\mathbf{y}_a$ is the component of $\mathbf{y}$ corresponding to the a-th variable, $\mathbf{y}_a^{label}$ is the component of $\mathbf{y}_a$ corresponding to the L labels, and $\mathbf{y}_a(p_{ai})$ is the component of $\mathbf{y}_a$ corresponding to the elements of p_{ai}.

Since Δ' is defined by linear equalities and inequalities, it is a polytope. With objective as maximum of linear functions and domain as a polytope, we have an LP relaxation of the corresponding MAP problem. □

Proposition 3. *In the limit $T \to 0$, the following problem for hierarchical Potts energies*

$$\min_{\mathbf{s} \in EP(F)} \quad g_T(\mathbf{s}) = \sum_{a=1}^{N} T \cdot \log \sum_{i=1}^{L} \exp(-\frac{s'_{ai}}{T}). \tag{24}$$

becomes:

$$-\min_{\mathbf{z} \in \Delta'} f(\mathbf{z}). \tag{25}$$

Proof. In the limit of $T \to 0$, we can rewrite the above problem as

$$\min_{\mathbf{s} \in EP(F)} \quad \sum_{a=1}^{N} \max_i(-s'_{ai}). \tag{26}$$

In vector form, the problem becomes

$$\min_{\mathbf{s} \in EP(F)} \max_{\mathbf{z} \in \Delta} -\langle \mathbf{z}, \mathbf{s}' \rangle \tag{27}$$

$$= - \max_{\mathbf{s} \in EP(F)} \min_{\mathbf{z} \in \Delta} \langle \mathbf{z}, \mathbf{s}' \rangle \tag{28}$$

$$\text{where } \Delta = \{\mathbf{z}_a \in \mathbb{R}^L | \mathbf{z}_a \succeq 0 \text{ and } \langle \mathbf{1}, \mathbf{z}_a \rangle = 1\} \tag{29}$$

where $\mathbf{z}_a$ is the component of $\mathbf{z}$ corresponding to the a-th variable. We can unpack $\mathbf{s}'$ using

$$s'_{ai} = \sum_{t \in p_{ai}} s_t. \tag{30}$$

and rewrite problem (28) as

$$-\max_{\mathbf{s} \in EP(F)} \min_{\mathbf{y} \in \Delta'} \langle \mathbf{y}, \mathbf{s} \rangle \tag{31}$$

The new constraint set Δ' ensures that the binary entries of labels and meta-labels is consistent:

$$\begin{aligned} \text{where } \Delta' = \{\mathbf{y}_a \in \mathbb{R}^M | \mathbf{y}_a \succeq 0,\ \langle \mathbf{1}, \mathbf{y}_a^{label} \rangle = 1, \\ \mathbf{y}_a(p_{ai}) = \mathbf{1} \text{ or } \mathbf{y}_a(p_{ai}) = \mathbf{0} \forall i \in [1, L]\} \end{aligned} \tag{32}$$

where $\mathbf{y}_a$ is the component of $\mathbf{y}$ corresponding to the a-th variable, $\mathbf{y}_a^{label}$ is the component of $\mathbf{y}_a$ corresponding to the L labels, and $\mathbf{y}_a(p_{ai})$ is the component of $\mathbf{y}_a$ corresponding to the elements of p_{ai}.

By the minimax theorem for LP, we can reorder the terms:

$$-\min_{\mathbf{y} \in \Delta'} \max_{\mathbf{s} \in EP(F)} \langle \mathbf{y}, \mathbf{s} \rangle \tag{33}$$

Recall that $\max_{\mathbf{s} \in EP(F)} \langle \mathbf{y}, \mathbf{s} \rangle$ is the value of the Lovasz extension of F at $\mathbf{y}$, that is, $f(\mathbf{y})$. Hence, as $T \to 0$, the marginal inference problem converts to minimising the Lovasz extension under the constraints Δ':

$$-\min_{\mathbf{y} \in \Delta'} f(\mathbf{y}). \tag{34}$$

□

Proposition 4. *The objective function $\widetilde{E}(\mathbf{z})$ of (T-LP-FULL) is the Lovasz extension of $F_{r-HST}(A) = \sum_{i=1}^{M} F_i(A)$, where*

$$\begin{aligned} F_i(A) = \sum_a \phi'_a(i)[|A \cap \{v_{ai}\}| = 1] + \\ \sum_{(a,b) \in \mathcal{N}} w_{ab} \cdot l_{T_i} \cdot [|A \cap \{v_{ai}, v_{bi}\}| = 1]. \end{aligned} \tag{35}$$

Proof. We observe that $F_{r-\text{HST}}$ is of exactly the same form as F_{Potts}, except that the Ising models F_i are defined over not just labels, but meta-labels as well. Using the same logic as in the proof of proposition 2, each F_i is the Lovasz extension of

$$\widetilde{E}_i(\mathbf{z}) = \left(\sum_{a \in \mathcal{X}} \phi'_a(i) \cdot z_{ai} + \sum_{(a,b) \in \mathcal{N}} w_{ab} \cdot l_{T_i} \cdot |z_{ai} - z_{bi}| \right) \tag{36}$$

and the results follows. □

Proposition 5. *Computing the subgradient of $E(\mathbf{y})$ in (P-LP) is equivalent to computing the conditional gradient for the submodular function F_{Potts}.*

Proof. Due to Edmond's greedy algorithm, the Lovasz extension f of a given submodular function can be written as

$$f(\mathbf{y}) = \max_{\mathbf{s} \in EP(F)} \mathbf{y}^T \cdot \mathbf{s} \tag{37}$$

Hence, $f(\mathbf{y})$ is the pointwise maximum of linear functions. The subdifferential of $f(\mathbf{y})$ at $\mathbf{y_0}$ is the differential of the 'active' linear function at $\mathbf{y_0}$. Hence,

$$\partial f(\mathbf{y}) = \operatorname*{argmax}_{\mathbf{s} \in EP(F)} \mathbf{y}^T \cdot \mathbf{s} \tag{38}$$

This is exactly the computation of conditional gradient, and hence we have proved the equivalence. □