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,

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

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

For $t = N + 1$,

$$\sum_{A^{N+1} \in \mathcal{M}^{N+1}} \exp(-s(A^{N+1})) = \sum_{i=1}^{L} \sum_{A^N \in \mathcal{M}^N} \exp(-s(A^N) - s_{N+1,i})$$

$$= \sum_{i=1}^{L} \exp(-s_{N+1,i}) \sum_{A^N \in \mathcal{M}^N} \exp(-s(A^N))$$

$$= \sum_{i=1}^{L} \exp(-s_{N+1,i}) \prod_{a=1}^{N} \sum_{i=1}^{L} \exp(-s_{ai})$$

$$= \prod_{a=1}^{N+1} \sum_{i=1}^{L} \exp(-s_{ai})$$  \hfill (3)

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

$$\min_{y \in \Delta} f(y).$$  \hfill (4)
Proof. By definition of a submodular extension and the Lovasz extension, \( E(x) = F(A_x) = f(1_{A_x}) \) for all valid labelings \( x \). Also, from property 1, \( f(y) \) is maximum of linear functions. Hence, \( f(y) \) is a piecewise linear relaxation of \( E(x) \).

The domain \( \Delta \) is a polytope formed by union of \( N \) probability simplices

\[
\Delta = \{ y_a \in \mathbb{R}^L | y_a \geq 0 \text{ and } \langle 1, y_a \rangle = 1 \} \tag{5}
\]

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

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

becomes

\[
-\min_{y \in \Delta} f(y). \tag{7}
\]

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

\[
\min_{s \in EP(F)} \sum_{a=1}^{N} \max_{i} (-s_{ai}). \tag{8}
\]

In vector form, the problem becomes

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

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

\( \Delta \) is the union of \( N \) probability simplices:

\[
\Delta = \{ y_a \in \mathbb{R}^L | y_a \geq 0 \text{ and } \langle 1, y_a \rangle = 1 \} \tag{11}
\]

where \( y_a \) is the component of \( y \) corresponding to the \( a \)-th variable. By the minimax theorem for LP, we can reorder the terms:

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

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

\[
-\min_{y \in \Delta} f(y) \tag{13}
\]
Proposition 2. The objective function $E(y)$ of the LP relaxation (P-LP) is the Lovasz extension of $F_{\text{Potts}}(A) = \sum_{i=1}^{L} F_i(A)$, where
\[
F_i(A) = \sum_{a} \phi_a(i)[|A \cap \{v_{ai}\}| = 1] + \sum_{(a,b) \in N} \frac{w_{ab}}{2} \cdot [|A \cap \{v_{ai}, v_{bi}\}| = 1].
\]

Proof. Since $F_{\text{Potts}}$ 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 $y = \{y_{ai}, y_{bi}\}$, such that $y_{ai}, y_{bi} \in [0, 1]$ and assume $y_{ai} > y_{bi}$. The Lovasz extension is:
\[
f(y) = y_{ai} \cdot [F_i(\{v_{ai}\}) - F_i(\{\})] + y_{bi} \cdot [F_i(\{v_{ai}, v_{bi}\}) - F_i(\{v_{ai}\})] \\
= y_{ai} \cdot [\phi_a(j) + \frac{w_{ab}}{2}] - y_{bi} \cdot [\phi_a(j) + \phi_b(j) - \frac{w_{ab}}{2}] \\
= \phi_a(j) \cdot y_{ai} + \phi_b(j) \cdot y_{bi} + \frac{w_{ab}}{2} \cdot (y_{ai} - y_{bi})
\]
(15)
In general for both orderings of $y_{ab}$ and $y_{bi}$, we can write
\[
f(y) = \phi_a(j) \cdot y_{ai} + \phi_b(j) \cdot y_{bi} + \frac{w_{ab}}{2} \cdot |y_{ai} - y_{bi}|
\]
(16)
Extending Lovasz extension (equation (16)) to all variables and labels gives $E(y)$ 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
\[
z_{ai} = \begin{cases} 
  y_{ai} & \text{if } i \in \mathcal{L} \\
  y_{a}(T_i) & \text{if } i \in \mathcal{R} - \mathcal{L}
\end{cases}
\]
(17)
We need to extend the definition of unary potentials to the expanded label space as follows:
\[
\phi'_a(i) = \begin{cases} 
  \phi_a(i) & \text{if } i \in \mathcal{L} \\
  0 & \text{if } i \in \mathcal{R} - \mathcal{L}
\end{cases}
\]
(18)
We can now rewrite problem (T-LP) in terms of new indicator variables $z_{ai}$:
\[
(T-LP-FULL) \quad \min \tilde{E}(z) = \sum_{i \in \mathcal{R}, a \in \mathcal{X}} \phi'_a(i) \cdot z_{ai} + \\
\sum_{i \in \mathcal{R}} \sum_{(a,b) \in \mathcal{N}} w_{ab} \cdot l_{T_i} \cdot |z_{ai} - z_{bi}|
\]
such that $z \in \Delta'$
(19)
where $\Delta'$ is the convex hull of the vectors satisfying
\[ \sum_{i \in \mathcal{L}} z_{ai} = 1, \quad z_{ai} \in \{0, 1\} \quad \forall a \in \mathcal{X}, \quad i \in \mathcal{L} \quad (20) \]
and
\[ z_{ai} = \sum_{j \in k(T_i)} z_{aj}, \quad \forall a \in \mathcal{X}, \quad i \in \mathcal{R} - \mathcal{L} \quad (21) \]
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
\[ \min_{z \in \Delta'} f(z). \quad (22) \]

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

We can write the domain $\Delta'$ as
\[ \Delta' = \{ y_a \in \mathbb{R}^M | y_a \geq 0, \quad \langle 1, y_a^{\text{label}} \rangle = 1, \quad y_a(p_{ai}) = 1 \text{ or } y_a(p_{ai}) = 0 \forall i \in [1, L] \} \quad (23) \]
where $y_a$ is the component of $y$ corresponding to the $a$-th variable, $y_a^{\text{label}}$ is the component of $y_a$ corresponding to the $L$ labels, and $y_a(p_{ai})$ is the component of $y_a$ corresponding to the elements of $p_{ai}$.

Since $\Delta'$ 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
\[ \min_{s \in EP(F)} g_T(s) = \sum_{a=1}^{N} T \cdot \log \sum_{i=1}^{L} \exp(-s'_{ai} T) \quad (24) \]
becomes:
\[ - \min_{z \in \Delta'} f(z). \quad (25) \]

**Proof.** In the limit of $T \to 0$, we can rewrite the above problem as
\[ \min_{s \in EP(F)} \sum_{a=1}^{N} \max_{i} (-s'_{ai}) \quad (26) \]
In vector form, the problem becomes
\[ \min_{s \in EP(F)} \max_{z} \langle z, s' \rangle \quad (27) \]
\[ = - \max_{s \in EP(F)} \min_{z} \langle z, s' \rangle \quad (28) \]
where $\Delta = \{ z_a \in \mathbb{R}^L | z_a \succeq 0 \text{ and } (1, z_a) = 1 \}$

(29)

where $z_a$ is the component of $z$ corresponding to the $a$-th variable. We can unpack $s'$ using

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

(30)

and rewrite problem (28) as

$$-\max_{s \in E\hat{P}(F)} \min_{y \in \Delta'} \langle y, s \rangle$$  

(31)

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

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

(32)

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

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

$$-\min_{y \in \Delta'} \max_{s \in E\hat{P}(F)} \langle y, s \rangle$$  

(33)

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

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

(34)

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

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

(35)

Proof. We observe that $F_{r-HST}$ is of exactly the same form as $F_{\text{Potts}}$, 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

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

(36)

and the results follows. \qed

Proposition 5. Computing the subgradient of $E(y)$ in (P-LP) is equivalent to computing the conditional gradient for the submodular function $F_{\text{Potts}}$. \hfill \qed
Proof. Due to Edmond’s greedy algorithm, the Lovasz extension $f$ of a given submodular function can be written as

$$f(y) = \max_{s \in EP(F)} y^T \cdot s$$  \hspace{1cm} (37)

Hence, $f(y)$ is the pointwise maximum of linear functions. The subdifferential of $f(y)$ at $y_0$ is the differential of the ‘active’ linear function at $y_0$. Hence,

$$\partial f(y) = \arg\max_{s \in EP(F)} y^T \cdot s$$  \hspace{1cm} (38)

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