Computing the fractional covering upper bound: Given potential functions $\phi_r(x_r)$, and nonnegative covering numbers $c_r$. Set $c_{p,r} = c_p/(c_r + \sum_{p' \in P(r)} c_{p'})$. for every initial values of messages $\lambda_{c \rightarrow p}(x_c)$:

1. For $t = 1, 2, ...$
   (a) For $r \in R$ do:
   \[
   \forall x_r, \forall p \in P(r) \quad \mu_{p \rightarrow r}(x_r) = c_p \log \left( \sum_{x_{p \setminus x_r}} \exp \left( \phi_p(x_p) + \sum_{c \in C(p) \setminus r} \lambda_{c \rightarrow r}(x_c) - \sum_{p' \in P(p)} \lambda_{p' \rightarrow r}(x_{p'}) / c_p \right) \right) \]
   \[
   \forall x_r, \forall p \in P(r) \quad \lambda_{r \rightarrow p}(x_r) = c_{p,r} \left( \phi_r(x_r) + \sum_{c \in C(r)} \lambda_{c \rightarrow r}(x_c) + \sum_{p' \in P(r)} \mu_{p' \rightarrow r}(x_{p'}) - \mu_{p \rightarrow r}(x_r) \right) \]

2. Output:
   (beliefs) $\forall r \in R \quad c_r \neq 0 : b^*_r(x_r) \propto \exp \left( \left( \phi_r(x_r) + \sum_{c \in C(r)} \lambda_{c \rightarrow r}(x_c) - \sum_{p \in P(r)} \lambda_{r \rightarrow p}(x_r) \right) / c_r \right)$
   \[
   \forall r \in R \quad c_r = 0 : \text{support}(b^*_r) \subset \text{argmax}_{x_r} \{ \phi_r(x_r) + \sum_{c \in C(r)} \lambda_{c \rightarrow r}(x_c) - \sum_{p \in P(r)} \lambda_{r \rightarrow p}(x_r) \} \]
   (bound) $\sum_{r, x_r} b^*_r(x_r)\phi_r(x_r) + \sum_r c_r H(b^*_r)$

Figure 1: The fractional covering upper bound appears in equation (2). The support of the beliefs are their non-zero entries, and when $c_r = 0$ it corresponds to the max-beliefs. When considering bipartite region graphs, this algorithm reduces to many of the previous message-passing algorithms, see Section 6.

3 Computing High-Order Upper Bounds

In the following we develop an efficient message-passing method to compute the region based upper bounds and their optimal beliefs $b_r(x_r)$, for fixed value of covering numbers $c_r$, as described in equation (2). These upper bounds depend on the non-negative fractional covering numbers, therefore correspond to maximizing a concave function subject to convex constraints. Such concave programs can be solved by minimizing their dual convex programs. Nevertheless, there are potentially many different convex dual programs, depending on the set of constraints, or Lagrange multipliers, one aims at satisfying. We realize that the probability simplex constraints for $b_r(x_r)$ are easier to satisfy, therefore we derive a dual program which ignores these constraints. For this purpose we use the entropy function as a barrier function over the probability simplex.

**Theorem 1.** Define the entropy as a barrier function over the probability simplex,
\[
H(b_r) = \left\{ \begin{array}{ll} -\sum_{x_r} b_r(x_r) \log b_r(x_r) & \text{if } b_r \in \Delta \\ -\infty & \text{otherwise} \end{array} \right. 
\]
where $b_r \in \Delta$ if $b_r(x_r) \geq 0$ and $\sum_{x_r} b_r(x_r) = 1$. The fractional covering numbers in equation (2) are non-negative, thus the bound is a concave function and its dual function takes the form
\[
D(\lambda) = \sum_r c_r \log \left( \sum_{x_r} \exp(\hat{\phi}_r(x_r)/c_r) \right) 
\]
where
\[
\hat{\phi}_r(x_r) = \phi_r(x_r) + \sum_{c \in C(r)} \lambda_{c \rightarrow r}(x_c) - \sum_{p \in P(r)} \lambda_{r \rightarrow p}(x_r) 
\]
In particular, strong duality holds and the primal optimal solution can be derived from the dual optimal solution
\[
b_r(x_r) \propto \exp \left( \hat{\phi}_r(x_r) / c_r \right) 
\]
Whenever $c_r = 0$ the corresponding primal optimal solution $b_r(x_r)$ corresponds to the max-beliefs, i.e., probability distributions over the maximal arguments of $\hat{\phi}_r(x_r)$.

**Proof:** Since we use the entropy as a barrier function over the probability simplex we only need to apply Lagrange multipliers $\lambda_{r \rightarrow p}(x_r)$ for the marginalization constraints $b_r(x_r) = \sum_{x_{p \setminus x_r}} b_p(x_p)$ for every region $r \in R$, every assignment $x_r$ and every parent $p \in P(r)$. Therefore the Lagrangian takes the form
\[
L(b_r, \lambda_{c \rightarrow p}) = \sum_{r, x_r} b_r(x_r)\phi_r(x_r) + \sum_r c_r H(b_r) \\
+ \sum_{r, x_r, p \in P(r)} \lambda_{r \rightarrow p}(x_r) \left( \sum_{x_p \setminus x_r} b_p(x_p) - b_r(x_r) \right) 
\]