Tightening the fractional covering upper bound: Given functions $\phi_i(r_i)$. For every initial values of $\lambda_i(r)$:

1. For $t=1,2,...$
   (a) For $r \in R$ do:
      \[
      \forall i \in r \quad \lambda_i(r) = \epsilon \log \left( \sum_{r_j \not= r} \exp \left( (\phi_i(r_i) + \lambda_i(r_i))/\epsilon \right) \right) - \phi_i(r) 
      \]
      additively normalize $\lambda_i(r)$ such that $\sum_{i \in r} \lambda_i(r) = 0$

2. Output: $\forall i$, set $q_i(r_i) \propto \exp \left( (\phi_i(r_i) + \lambda_i(r_i))/\epsilon \right)$ and $\forall r$ set $c_r = q_i(r)$, for any of $i \in r$.

Figure 2: The dual decomposition algorithm for recovering the direction of descent, described in equation (4). This algorithm can be seen as message-passing by reformulating the indexes $\lambda_i(r) \leftrightarrow \lambda_i \rightarrow r$. This demonstrates the differences between our two algorithms. The similarity between the two algorithms is a consequence of the block dual steps and the use of the entropy barrier method.

For computational efficiency we solve the above linear program by dual block ascent. Unfortunately, dual coordinate ascent may be sub-optimal when the dual program is not smooth, or equivalently the primal program is not strictly convex. Therefore we use the entropy barrier method to make the primal program strictly convex, which results in a smooth dual program. The following theorem shows that using the barrier method does not affect the quality of the solution.

**Theorem 6.** Let $\phi_i(r_i) = H(b^*_i)/|r_i|$ and consider the linear program for finding the direction of descent

\[
\min_{\forall i \quad q_i \text{ is probability}} \sum_{i,r_i} q_i(r_i)\phi_i(r_i)
\]

Then the primal-dual programs

\[
\begin{align*}
\text{(primal)} & \quad \min_{\forall i \quad q_i \text{ is probability}} \sum_{i,r_i} q_i(r_i)\phi_i(r_i) - \epsilon \sum_i H(q_i) \\
\text{(dual)} & \quad \max_{\sum_i \lambda_i(r_i)=0} \epsilon \sum_i \log \left( \sum_{r_i} \exp \left( (\phi_i(r_i) + \lambda_i(r_i))/\epsilon \right) \right)
\end{align*}
\]

are $\delta-$approximations of the original linear program, where $\delta = \sum_i \epsilon |\log |r_i||$

**Proof:** The entropy $H(q_i)$ is a non-negative measure, thus the primal program lower bounds the original linear program. The bound holds since $H(q_i) \leq |\log |r_i||$. The dual program is also a $\delta-$approximation since strong duality holds. \[ \square \]

Having a smooth dual program which relates to a strictly convex primal program, we can perform dual...