Algorithm 11.2 Convergent message passing for Bethe cluster graph with convex counting numbers

Procedure Convex-BP-Msg (
  $\psi_r(\boldsymbol{C}_r)$ // set of initial potentials
  $\sigma_{i\to r}(\boldsymbol{C}_r)$ // Current node-to-region messages
)
  for $i = 1, \ldots, n$
    // Compute incoming messages from neighboring regions to X_i
    for $r \in \mathrm{Nb}_i$
      $\delta_{r\to i}(X_i) \leftarrow \sum_{\boldsymbol{C}_r - X_i} \left(\psi_r(\boldsymbol{C}_r) \prod_{j \in \mathrm{Nb}_r - \{i\}} \sigma_{j\to r}(\boldsymbol{C}_r)\right)^{\frac{1}{\hat{\nu}_{i,r}}}$
    // Compute beliefs for X_i, renormalizing to avoid numerical underflows
    $\beta_i(X_i) \leftarrow \frac{1}{Z_{X_i}} \prod_{r \in \mathrm{Nb}_i} (\delta_{r\to i}(X_i))^{\hat{\nu}_{i,r}/\hat{\nu}_i}$
    // Compute outgoing messages from X_i to neighboring regions
    for $r \in \mathrm{Nb}_i$
      $\sigma_{i\to r}(\boldsymbol{C}_r) \leftarrow \left(\psi_r(\boldsymbol{C}_r) \prod_{j \in \mathrm{Nb}_r - \{i\}} \sigma_{j\to r}(\boldsymbol{C}_r)\right)^{-\frac{\nu_{i,r}}{\hat{\nu}_{i,r}}} \left(\frac{\beta_i(X_i)}{\delta_{r\to i}(X_i)}\right)^{\nu_r}$
  return $\{\sigma_{i\to r}(\boldsymbol{C}_r)\}_{i, r \in \mathrm{Nb}_i}$

factor graph

and outgoing messages $\sigma_{i\to r}(\boldsymbol{C}_r)$ from variables to regions (essentially passing messages over the *factor graph*). The overall process is initialized (in the first message passing iteration) by setting $\sigma_{i\to r} = 1$. This algorithm is guaranteed to converge to the global maximum of our convex energy functional.

This derivation applies to any set of convex counting numbers, leaving open the question of which counting numbers are likely to be give the best approximation. Although there is currently no theoretical analysis answering this question, intuitively, we might argue that we want the counting numbers for different regions to be as close as possible to uniform. This intuition is also supported by the fact that the Bethe approximation, which sets all $\kappa_r = 1$, obtains very high-quality approximations when it converges. Thus, we can try to select nonnegative coefficients ν_i, ν_r, and $\nu_{i,r}$ for which κ_r and κ_i, defined via equation (11.22), satisfy equation (11.21) and minimize

$$\sum_{r \in \mathbf{R}^+} (\kappa_r - 1)^2. \tag{11.25}$$

TRW

Other choices are also possible. For example, the *tree-reweighted belief propagation (TRW)* algorithm computes convex counting numbers for a pairwise Markov network using the following process: We first define a probability distribution ρ over trees $\mathcal{T}$ in the network, such that each edge in the pairwise network is present in at least one tree. This distribution defines a set of weights:

$$\begin{array}{ll} \kappa_i & = -\sum_{\mathcal{T} \ni X_i} \rho(\mathcal{T}) \\ \kappa_{i,j} & = \sum_{\mathcal{T} \ni (X_i, X_j)} \rho(\mathcal{T}) \end{array} \tag{11.26}$$