Relax then Compensate: On Max-Product Belief Propagation and More (Appendix)

Arthur Choi
Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095
aychoi@cs.ucla.edu

Adnan Darwiche
Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095
darwiche@cs.ucla.edu

First, we restate Equations and Theorems from the main paper.

Consider the REC-BP condition for a single equivalence constraint relaxed:

$$\mathsf{c\text{-}map}(X_i\!=\!x) \quad = \quad \mathsf{c\text{-}map}(X_j\!=\!x) \quad = \quad \theta(X_i\!=\!x) + \theta(X_j\!=\!x) + \gamma \tag{1}$$

Proposition 1 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing an equivalence constraint* $X_i \equiv X_j$ *that split* $\mathcal{M}$ *into two independent sub-models. Then the compensation has parameters satisfying Equation 1 iff* $\mathsf{c\text{-}map}(X_i\!=\!x) = \mathsf{c\text{-}map}(X_j\!=\!x) = \mathsf{map}(X_i\!=\!x, X_j\!=\!x) + \gamma$.

Consider the REC-I condition for a single equivalence constraint relaxed:

$$\mathsf{c\text{-}map}(X_i\!=\!x) \quad = \quad \mathsf{c\text{-}map}(X_j\!=\!x) \quad = \quad 2 \cdot [\theta(X_i\!=\!x) + \theta(X_j\!=\!x)] \tag{2}$$

Proposition 2 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing an equivalence constraint* $X_i \equiv X_j$ *in* $\mathcal{M}$*. If* $\mathsf{c\text{-}map}(.)$ *has valid configurations and scaled values, then* $\mathsf{c\text{-}map}(.)$ *satisfies Equation 2.*

Consider the REC-BP condition for a k equivalence constraints relaxed:

$$\mathsf{c\text{-}map}(X_i\!=\!x) \quad = \quad \mathsf{c\text{-}map}(X_j\!=\!x) \quad = \quad \theta(X_i\!=\!x) + \theta(X_j\!=\!x) + \gamma \tag{3}$$

Theorem 1 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing enough equivalence constraints* $X_i \equiv X_j$ *in* $\mathcal{M}$ *to render it fully disconnected. Then a compensation whose parameters satisfy Equation 3 has values* $\exp\{\mathsf{c\text{-}map}(X_i\!=\!x)\}$ *that correspond to the max-marginals of a fixed-point of max-product belief propagation run on* $\mathcal{M}$*, and vice-versa.*

Consider REC-I condition for a k equivalence constraints relaxed:

$$\mathsf{c\text{-}map}(X_i\!=\!x) \quad = \quad \mathsf{c\text{-}map}(X_j\!=\!x) \quad = \quad (1+k)[\theta(X_i\!=\!x) + \theta(X_j\!=\!x)] \tag{4}$$

Theorem 2 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing* k *equivalence constraints* $X_i \equiv X_j$*. If the compensation has parameters satisfying either Eqs. 3 or 4, and if* $\mathbf{x}^\star$ *is an optimal assignment for the compensation that is also valid, then: (1)* $\mathbf{x}^\star$ *is optimal for the model* $\mathcal{M}$*, and (2)* $\frac{1}{1+k}\mathsf{c\text{-}map}^\star = \mathsf{map}^\star$.

Theorem 3 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing* k *equivalence constraints* $X_i \equiv X_j$*. If the compensation has parameters satisfying Equation 4, then* $\mathsf{map}^\star \leq \frac{1}{1+k}\mathsf{c\text{-}map}^\star$.

Theorem 4 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing* k *equivalence constraints* $X_i \equiv X_j$*. If the compensation has parameters satisfying Eq. 4, and if* $\mathbf{z}$ *is a partial assignment that sets the same sign to variables* X_i *and* X_j*, for any equivalence constraint* $X_i \equiv X_j$ *deleted, then:* $\mathsf{map}(\mathbf{z}) \leq \frac{1}{1+k}\mathsf{c\text{-}map}(\mathbf{z})$.

A Proofs

Proof of Proposition 1 Let $\mathsf{c\text{-}map}_i(X_i)$ and $\mathsf{c\text{-}map}_j(X_j)$ denote a compensation's MAP values in subnetworks $\mathcal{M}_i$ and $\mathcal{M}_j$. Then:

$$\mathsf{c\text{-}map}_i(X_i) = \mathsf{r\text{-}map}_i(X_i) + \theta(X_i) \quad \text{and} \quad \mathsf{c\text{-}map}_j(X_j) = \mathsf{r\text{-}map}_j(X_j) + \theta(X_j).$$

Using Equation 1, we find

$$\begin{aligned}
&\mathsf{c\text{-}map}(X_i = x) = \mathsf{c\text{-}map}_i(X_i = x) + \mathsf{c\text{-}map}_j^\star \\
&\quad = \quad \mathsf{r\text{-}map}_i(X_i = x) + \theta(X_i = x) + \mathsf{c\text{-}map}_j^\star = \theta(X_i = x) + \theta(X_j = x) + \gamma \\
&\mathsf{c\text{-}map}(X_j = x) = \mathsf{c\text{-}map}_i^\star + \mathsf{c\text{-}map}_j(X_j = x) \\
&\quad = \quad \mathsf{c\text{-}map}_i^\star + \mathsf{r\text{-}map}_j(X_j = x) + \theta(X_j = x) = \theta(X_i = x) + \theta(X_j = x) + \gamma
\end{aligned}$$

thus we have that

$$\theta(X_j = x) = \mathsf{r\text{-}map}_i(X_i = x) + \mathsf{c\text{-}map}_j^\star - \gamma \quad \text{and} \quad \theta(X_i = x) = \mathsf{c\text{-}map}_i^\star + \mathsf{r\text{-}map}_j(X_j = x) - \gamma.$$

Substituting back in, we find that

$$\begin{aligned}
&\mathsf{c\text{-}map}(X_i = x) = \mathsf{c\text{-}map}(X_j = x) \\
&\quad = \quad \mathsf{r\text{-}map}_i(X_i = x) + \mathsf{r\text{-}map}_j(X_j = x) + \mathsf{c\text{-}map}_1^\star + \mathsf{c\text{-}map}_2^\star - \gamma = \mathsf{map}(X_i = x, X_j = x) + \gamma
\end{aligned}$$

since $\mathsf{r\text{-}map}_i(X_i = x) + \mathsf{r\text{-}map}_j(X_j = x) = \mathsf{map}(X_i = x, X_j = x)$ and $\mathsf{c\text{-}map}_1^\star + \mathsf{c\text{-}map}_2^\star = \mathsf{c\text{-}map}^\star$ and $\gamma = \frac{1}{2}\mathsf{c\text{-}map}^\star$. We can reverse the steps, for the other direction. □

Proof of Proposition 2 It suffices to show $\theta(X_i = x) + \theta(X_j = x) = \frac{\kappa - 1}{\kappa} \cdot \mathsf{c\text{-}map}(X_i = x, X_j = x)$. Now,

$$\begin{aligned}
\mathsf{c\text{-}map}(X_i = x, X_j = x) &= \mathsf{map}(X_i = x, X_j = x) + [\theta(X_i = x) + \theta(X_j = x)] \\
&= \kappa^{-1} \cdot \mathsf{c\text{-}map}(X_i = x, X_j = x) + [\theta(X_i = x) + \theta(X_j = x)]
\end{aligned}$$

so $\theta(X_i = x) + \theta(X_j = x) = (1 - \frac{1}{\kappa}) \cdot \mathsf{c\text{-}map}(X_i = x, X_j = x) = \frac{\kappa - 1}{\kappa} \cdot \mathsf{c\text{-}map}(X_i = x, X_j = x)$. □

Proof of Theorem 1 Analogous to the correspondence of the ED-BP algorithm and sum-product belief propagation, shown in [1]. □

Lemma 1 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing* k *equivalence constraints* $X_i \equiv X_j$*. If the compensation has parameters satisfying Equation 4, and if* $\tilde{\mathbf{x}}$ *is a complete assignment that is also valid, then* $\mathsf{map}(\tilde{\mathbf{x}}) \leq \frac{1}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{x}})$*, with equality if* $\tilde{\mathbf{x}}$ *is also optimal for* $\mathsf{c\text{-}map}(.)$.

Proof First, we have

$$\mathsf{c\text{-}map}(\tilde{\mathbf{x}}) = \mathsf{map}(\tilde{\mathbf{x}}) + \textstyle\sum_{X_i \equiv X_j} [\theta(X_i = x) + \theta(X_j = x)] = \mathsf{map}(\tilde{\mathbf{x}}) + \sum_{X_i \equiv X_j} \frac{1}{1+k}\mathsf{c\text{-}map}(X_i = x).$$

Note that x is the state assumed by X_i and X_j in assignment $\tilde{\mathbf{x}}$. Since $\mathsf{c\text{-}map}(X_i = x) \geq \mathsf{c\text{-}map}(\tilde{\mathbf{x}})$,

$$\mathsf{c\text{-}map}(\tilde{\mathbf{x}}) \geq \mathsf{map}(\tilde{\mathbf{x}}) + \textstyle\sum_{X_i \equiv X_j} \frac{1}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{x}}) = \mathsf{map}(\tilde{\mathbf{x}}) + \frac{k}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{x}})$$

and thus $\frac{1}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{x}}) \geq \mathsf{map}(\tilde{\mathbf{x}})$. In the case where $\mathsf{c\text{-}map}(\tilde{\mathbf{x}}) = \mathsf{c\text{-}map}^\star$, we have that $\mathsf{c\text{-}map}(X_i = x) = \mathsf{c\text{-}map}(\tilde{\mathbf{x}})$ for all $X_i \equiv X_j$, so we have $\mathsf{map}(\tilde{\mathbf{x}}) = \frac{1}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{x}})$. □

Proof of Theorem 2 Note first in a REC-BP compensation, from Equation 3, that

$$\theta(X_i = x) + \theta(X_j = x) = \mathsf{c\text{-}map}(X_i = x) - \frac{k}{1+k}\mathsf{c\text{-}map}^\star = \mathsf{c\text{-}map}^\star - \frac{k}{1+k}\mathsf{c\text{-}map}^\star = \frac{1}{1+k}\mathsf{c\text{-}map}^\star$$

When we decompose $\mathsf{c\text{-}map}(\mathbf{x}^\star)$ into the original factors, i.e., $\mathsf{map}(\mathbf{x}^\star)$, and the auxiliary parameters $\theta(X_i = x) + \theta(X_j = x)$:

$$\begin{aligned}
\mathsf{c\text{-}map}(\mathbf{x}^\star) &= \mathsf{map}(\mathbf{x}^\star) + \sum_{X_i \equiv X_j} [\theta(X_i = x) + \theta(X_j = x)] \\
&= \mathsf{map}(\mathbf{x}^\star) + \textstyle\sum_{X_i \equiv X_j} \frac{1}{1+k}\mathsf{c\text{-}map}^\star = \mathsf{map}(\mathbf{x}^\star) + \frac{k}{1+k}\mathsf{c\text{-}map}^\star
\end{aligned}$$

and $\frac{1}{1+k}\mathsf{c\text{-}map}(\mathbf{x}^\star) = \mathsf{map}(\mathbf{x}^\star)$. Similarly, for a REC-I compensation, since $\theta(X_i\!=\!x) + \theta(X_j\!=\!x) = \frac{1}{1+k}\mathsf{c\text{-}map}(X_i\!=\!x) = \frac{1}{1+k}\mathsf{c\text{-}map}^\star$.

We now want to show that $\mathbf{x}^\star$ must be optimal for $\mathsf{map}(.)$. Suppose for contradiction that there is an assignment $\hat{\mathbf{x}}$ such that $\mathsf{map}(\hat{\mathbf{x}}) > \mathsf{map}(\mathbf{x}^\star)$. Then:

$$\begin{aligned}
\mathsf{c\text{-}map}(\hat{\mathbf{x}}) &= \mathsf{map}(\hat{\mathbf{x}}) + \sum_{X_i \equiv X_j} [\theta(X_i\!=\!x) + \theta(X_j\!=\!x)] \\
&= \mathsf{map}(\hat{\mathbf{x}}) + \textstyle\sum_{X_i \equiv X_j} [\mathsf{c\text{-}map}(X_i\!=\!x) - \frac{k}{1+k}\mathsf{c\text{-}map}^\star] \\
&\geq \mathsf{map}(\hat{\mathbf{x}}) + \textstyle\sum_{X_i \equiv X_j} [\mathsf{c\text{-}map}(\hat{\mathbf{x}}) - \frac{k}{1+k}\mathsf{c\text{-}map}^\star] \\
&= \mathsf{map}(\hat{\mathbf{x}}) + k \cdot \mathsf{c\text{-}map}(\hat{\mathbf{x}}) - \tfrac{k^2}{1+k}\mathsf{c\text{-}map}^\star \\
&> \mathsf{map}(\mathbf{x}^\star) + k \cdot \mathsf{c\text{-}map}(\hat{\mathbf{x}}) - \tfrac{k^2}{1+k}\mathsf{c\text{-}map}^\star \\
&= \tfrac{1}{1+k}\mathsf{c\text{-}map}^\star + k \cdot \mathsf{c\text{-}map}(\hat{\mathbf{x}}) - \tfrac{k^2}{1+k}\mathsf{c\text{-}map}^\star \\
&= k \cdot \mathsf{c\text{-}map}(\hat{\mathbf{x}}) + \tfrac{1-k^2}{1+k}\mathsf{c\text{-}map}^\star = k \cdot \mathsf{c\text{-}map}(\hat{\mathbf{x}}) + (1-k) \cdot \mathsf{c\text{-}map}^\star
\end{aligned}$$

and thus $\mathsf{c\text{-}map}(\hat{\mathbf{x}}) > \mathsf{c\text{-}map}^\star$ which contradicts the optimality of $\mathsf{c\text{-}map}^\star$.

We know for a REC-I compensation, that $\mathbf{x}^\star$ must be optimal for $\mathsf{map}(.)$. Otherwise it would not be optimal for $\mathsf{c\text{-}map}(.)$, by Lemma 1. $\square$

Proof of Theorem 3 Let $\mathbf{x}^\star$ be an optimal assignment for $\mathsf{map}(.)$. Since $\mathbf{x}^\star$ must also be valid, we have by Lemma 1 that $\mathsf{map}^\star = \mathsf{map}(\mathbf{x}^\star) \leq \frac{1}{1+k}\mathsf{c\text{-}map}(\mathbf{x}^\star) \leq \frac{1}{1+k}\mathsf{c\text{-}map}^\star$. $\square$

Proof of Theorem 4 We have that

$$\begin{aligned}
\mathsf{map}(\tilde{\mathbf{z}}) &= \max_{\tilde{\mathbf{x}} \sim \tilde{\mathbf{z}}} \mathsf{map}(\tilde{\mathbf{x}}) \leq \max_{\tilde{\mathbf{x}} \sim \tilde{\mathbf{z}}} \tfrac{1}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{x}}) \\
&\leq \max_{\mathbf{x} \sim \tilde{\mathbf{z}}} \tfrac{1}{1+k}\mathsf{c\text{-}map}(\mathbf{x}) = \tfrac{1}{1+k}\mathsf{c\text{-}map}(\tilde{\mathbf{z}})
\end{aligned}$$

where the first inequality follows from Lemma 1. $\square$

Proposition 3 *Let* $\mathsf{map}(.)$ *denote the MAP values of a model* $\mathcal{M}$*, and let* $\mathsf{c\text{-}map}(.)$ *denote the MAP values of a compensation that results from relaxing a single equivalence constraint* $X_i \equiv X_j$*. If the compensation has parameters satisfying either Eqs. 1 or 2, and if* $\mathbf{x}^\star$ *is an optimal assignment for the compensation that is also valid, then:*

$$\mathsf{map}^\star \leq \frac{1}{2}\mathsf{c\text{-}map}^\star \leq \mathsf{r\text{-}map}^\star.$$

Proof The inequality $\mathsf{map}^\star \leq \frac{1}{2}\mathsf{c\text{-}map}^\star$ is implied by Theorem 3. The inequality $\frac{1}{2}\mathsf{c\text{-}map}^\star \leq \mathsf{r\text{-}map}^\star$, follows trivially if $\mathsf{c\text{-}map}^\star = \mathsf{c\text{-}map}(X_i\!=\!x, X_j\!=\!x)$, for some state x, since $\frac{1}{2}\mathsf{c\text{-}map}^\star = \mathsf{map}^\star$ in this case, and $\mathsf{map}^\star \leq \mathsf{r\text{-}map}^\star$. Then assume that if $\mathsf{c\text{-}map}^\star = \mathsf{c\text{-}map}(X_i\!=\!x_i, X_j\!=\!x_j)$, then $x_i \neq x_j$. It must then be that $\mathsf{c\text{-}map}(X_i\!=\!x_i) = \mathsf{c\text{-}map}(X_i\!=\!x_i, X_j\!=\!x_j) = \mathsf{c\text{-}map}(X_j\!=\!x_j)$, so there are at least 2 states x where the optimal MAP value $\mathsf{c\text{-}map}^\star$ is achieved. Select s pairs of optimal assignments $X_i\!=\!x_i, X_j\!=\!x_j$ where X_i and X_j are set to a particular state x exactly once (like $X_i\!=\!x_1, X_j\!=\!x_2$ and $X_i\!=\!x_2, X_j\!=\!x_3$ and $X_i\!=\!x_3, X_j\!=\!x_1$). We know such a selection exists. Suppose, without loss of generality, that $X_i\!=\!x_1, X_j\!=\!x_2$ is an optimal assignment (we can relabel states). Then there must be an optimal assignment $X_i\!=\!x_2, X_j\!=\!x$ for some state x, since $\mathsf{c\text{-}map}(X_i\!=\!x_2) = \mathsf{c\text{-}map}(X_j\!=\!x_2)$. If $x = x_1$ then we are done. Otherwise, we can assume w.l.o.g. that $x = x_3$, and we can repeat the above reasoning.

Using these s assignments $X_i = x_i, X_j = x_j$, we have that:

$$\begin{aligned}
s \cdot \textsf{c-map}^\star &= \sum_{X_i = x_i, X_j = x_j} \textsf{c-map}(X_i = x_i, X_j = x_j) \\
&= \sum_{X_i = x_i, X_j = x_j} \textsf{r-map}(X_i = x_i, X_j = x_j) + \theta(X_i = x_i) + \theta(X_j = x_j) \\
&= \Big[\sum_{X_i = x_i, X_j = x_j} \textsf{r-map}(X_i = x_i, X_j = x_j) \Big] + \Big[\sum_{x_i = x_j} \theta(X_i = x_i) + \theta(X_j = x_j) \Big] \\
&= \Big[\sum_{X_i = x_i, X_j = x_j} \textsf{r-map}(X_i = x_i, X_j = x_j) \Big] + \Big[\sum_{x_i = x_j} \frac{1}{2} \textsf{c-map}(X_i = x_i) \Big] \\
&= \Big[\sum_{X_i = x_i, X_j = x_j} \textsf{r-map}(X_i = x_i, X_j = x_j) \Big] + \Big[\sum_{x_i = x_j} \frac{1}{2} \textsf{c-map}^\star \Big] \\
&= \Big[\sum_{X_i = x_i, X_j = x_j} \textsf{r-map}(X_i = x_i, X_j = x_j) \Big] + \frac{s}{2} \textsf{c-map}^\star
\end{aligned}$$

and thus

$$\frac{1}{2} \textsf{c-map}^\star = \frac{1}{s} \sum_{X_i = x_i, X_j = x_j} \textsf{r-map}(X_i = x_i, X_j = x_j) \leq \frac{1}{s} \sum_{X_i = x_i, X_j = x_j} \textsf{r-map}^\star = \textsf{r-map}^\star$$

as desired. □

References

[1] Arthur Choi and Adnan Darwiche. An edge deletion semantics for belief propagation and its practical impact on approximation quality. In *AAAI*, pages 1107–1114, 2006.