Supplementary Material to "Cooled and Relaxed Survey Propagation for MRFs"

Hai Leong Chieu[1,2], **Wee Sun Lee**[2]
[1]Singapore MIT Alliance
[2]Department of Computer Science
National University of Singapore
haileong@nus.edu.sg,leews@comp.nus.edu.sg

Yee-Whye Teh
Gatsby Computational Neuroscience Unit
University College London
ywteh@gatsby.ucl.ac.uk

Abstract

This is the supplementary material for the submission to NIPS 2007, entitled "Cooled and Relaxed Survey Propagation for MRFs". The purpose of this material is to prove the update equations of Relaxed Survey Propagation (RSP) in the main paper.

1 Introduction

In this paper, we will refer to our submission to NIPS 2007 entitled "Cooled and Relaxed Survey Propagation for MRFs" as the "main paper". The objective of this paper is to derive the RSP update equations given in the main paper. In Section 2, we will derive general RSP update equations for weighted MAX-SAT (WMS) problems. In Section 3, we show that under the settings of Section 3.1 in the main paper, the update equations can be simplified to the simple equations given in Section 3.3 in the main paper.

2 RSP for weighted MAX-SAT problems

In this section, we derive the update equations of Relaxed Survey Propagation (RSP) for a weighted MAX-SAT (WMS) problem, $W = (B, C)$. The variables in B are binary, but in the survey propagation framework, we allow the variables to take values in $\{0, 1, *\}$. The value * is the joker state, and variables taking the value * is free to take either 0 or 1, without violating any clauses.

For $\alpha \in C$, define $u_{\alpha,j}$ (resp. $s_{\alpha,j}$) as the value of $\sigma_j \in \{0, 1\}$ that violates (resp. satisfies) clause α. Let $C(\alpha)$ be the set of variables in clause α.

$$\begin{array}{rcl} C(j) & = & \{\alpha \in C : j \in C(\alpha)\} \\ C^+(j) & = & \{\alpha \in C(j); s_{\alpha,j} = 1\} \\ C^-(j) & = & \{\alpha \in C(j); s_{\alpha,j} = 0\} \\ C^s_\alpha(j) & = & \{\beta \in C(j) \setminus \{\alpha\}; s_{\alpha,j} = s_{\beta,j}\} \\ C^u_\alpha(j) & = & \{\beta \in C(j) \setminus \{\alpha\}; s_{\alpha,j} \neq s_{\beta,j}\} \end{array} \tag{1}$$

In the above definitions, $C(j)$ is the set of clauses containing σ_j, $C^+(j)$ (resp. $C^-(j)$) is the set of clauses that contains σ_j as a positive literal (resp. negative literal). $C^s_\alpha(j)$ (resp. $C^u_\alpha(j)$ is the set of clauses containing σ_j that agrees (resp. disagrees) with the clause α concerning the variable σ_j.

RSP is the sum product belief algorithm applied to the *relaxed* MRF defined in the main paper. The RSP update equations can be derived in a similar manner as the SP-ρ algorithm in [2]. Each message from a clause α to a variable k is a vector of length $x \times |P(k)|$, where $P(k)$ is the set of all possible parents sets of the variable k. Due to symmetries in the variable and clause compatibilities, these messages can be grouped as follows (refer to [2] for more detailed explanations):

$$M_{\alpha\to k}(x_k, P_k) = \begin{cases} M^s_{\alpha\to k} & \text{if } x_k = s_{a,k}, P_k = S \cup \{\alpha\} \text{ for some } S \subseteq C^s_\alpha(k) \\ M^u_{\alpha\to k} & \text{if } x_k = u_{a,k}, P_k \subseteq C^u_a(k) \\ M^*_{\alpha\to k} & \text{if } x_k = s_{a,k}, P_k \subseteq C^s_\alpha(k), \alpha \notin P_k \text{ or } x_k = *, P_k = \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Similarly, the variable to clause messages $M_{k\to a}$ can be grouped as follows [2]:

$$\begin{aligned} R^s_{k\to\alpha} &= \textstyle\sum_{S\subseteq C^s_\alpha(k)} M_{k\to\alpha}(s_{\alpha,k}, S\cup\{\alpha\}) \\ R^u_{k\to\alpha} &= \textstyle\sum_{P_k\subseteq C^u_\alpha(k)} M_{k\to\alpha}(u_{\alpha,k}, P_k) \\ R^*_{k\to\alpha} &= \textstyle\sum_{P_k\subseteq C^s_\alpha(k)} M_{k\to\alpha}(s_{\alpha,k}, P_k) + M_{k\to\alpha}(*, \emptyset) \end{aligned} \tag{3}$$

With these definitions, the update equations are shown in Figure 1. Note that the equations in Figure 1 are similar to those for SP-ρ, given in [2]. The difference between SP-ρ and RSP is that for the message $M^{s,u,*}_{\alpha\to k}$, satisfied clauses have weights, and violated clauses are allowed (with a penalty). Hence in the equations 4, 5 and 6, the multiplicative factors $\exp(w_\alpha)$ are the weights, and the factors $\exp(y_{\mathrm{src}(\alpha)})$ are the penalties. The equations in Figure 1 will be proportional to those for SP-ρ if all the y's are taken to infinity. In this case, the factor $\exp(w_\alpha)$ will be a constant factor.

$$M^s_{\alpha\to i} = e^{w_\alpha}\left[\prod_{j\in C(\alpha)-\{i\}} R^u_{j\to\alpha}\right] \tag{4}$$

$$\begin{aligned} M^u_{\alpha\to i} = {} & e^{w_\alpha}\left[\prod_{j\in C(\alpha)-\{i\}} (R^u_{j\to\alpha} + R^*_{j\to\alpha}) + \sum_{k\in C(\alpha)-\{i\}} (R^s_{k\to\alpha} - R^*_{k\to\alpha}) \prod_{j\in C(\alpha)-\{i,k\}} R^u_{j\to\alpha}\right] \\ & + (e^{-y_{\mathrm{src}(\alpha)}} - e^{w_\alpha}) \prod_{j\in C(\alpha)-\{i\}} R^u_{j\to\alpha} \end{aligned} \tag{5}$$

$$M^*_{\alpha\to i} = e^{w_\alpha}\left[\prod_{j\in C(\alpha)-\{i\}} (R^u_{j\to\alpha} + R^*_{j\to\alpha}) - \prod_{j\in C(\alpha)-\{i\}} R^u_{j\to\alpha}\right] \tag{6}$$

$$R^s_{i\to\alpha} = \prod_{\beta\in C^u_\alpha(i)} M^u_{\beta\to i}\left[\prod_{\beta\in C^s_\alpha(i)} (M^s_{\beta\to i} + M^*_{\beta\to i})\right] \tag{7}$$

$$R^u_{i\to\alpha} = \prod_{\beta\in C^s_\alpha(i)} M^u_{\beta\to i}\left[\prod_{\beta\in C^u_\alpha(i)} (M^s_{\beta\to i} + M^*_{\beta\to i}) - (1-\omega_0) \prod_{\beta\in C^u_\alpha(i)} M^*_{\beta\to i}\right] \tag{8}$$

$$\begin{aligned} R^*_{i\to\alpha} = {} & \prod_{\beta\in C^u_\alpha(i)} M^u_{\beta\to i}\left[\prod_{\beta\in C^s_\alpha(i)} (M^s_{\beta\to i} + M^*_{\beta\to i}) - (1-\omega_0) \prod_{\beta\in C^s_\alpha(i)} M^*_{\beta\to i}\right] \\ & + \omega_* \prod_{\beta\in C^s_\alpha(i)\cup C^u_\alpha(i)} M^*_{\beta\to i} \end{aligned} \tag{9}$$

$$B_i(0) \propto \prod_{\beta\in C^+(i)} M^u_{\beta\to i}\left[\prod_{\beta\in C^-(i)} (M^s_{\beta\to i} + M^*_{\beta\to i}) - \omega_* \prod_{\beta\in C^-(i)} M^*_{\beta\to i}\right] \tag{10}$$

$$B_i(1) \propto \prod_{\beta\in C^-(i)} M^u_{\beta\to i}\left[\prod_{\beta\in C^+(i)} (M^s_{\beta\to i} + M^*_{\beta\to i}) - \omega_* \prod_{\beta\in C^+(i)} M^*_{\beta\to i}\right] \tag{11}$$

$$B_i(*) \propto \prod_{\beta\in C(i)} M^*_{\beta\to i} \tag{12}$$

Figure 1: The update equations for RSP.

3 RSP for Markov Random Fields

In this section, we derive the update equations in the main paper, for the MRF $G_s = (V_s, F_s)$.

Definition 1. *As in Section 3.1 of the main paper, we consider RSP under the following settings:*

1. *Set* $\omega_* = 1$ *and* $\omega_0 = 0$.
2. *For positivity clauses* $\beta(i)$*, let* $y_i = 0$.
3. *Without lost of generality, we assume that in the original MRF* $G = (V, F)$*, single-variable factors are defined on all variables.*

Under these settings, we prove in the main paper that the joint distribution on the *relaxed* MRF is approximately equal to that on the original MRF, and that RSP estimates marginals on the original MRF.

Recall that variables in V_s are of the form $\lambda_{(i,x_i)} = (\sigma_{(i,x_i)}, P_{(i,x_i)})$, where $X_i \in V$ are variables in the MRF $G = (V, F)$, and x_i are their values. By Lemma 2 in the main paper, there is a one-to-one mapping between $\lambda_{(i,x_i)}$ and $\sigma_{(i,x_i)}$.

We consider the messages $M_{\gamma\to(i,x_i)}$ for two separate cases: (1) $\gamma = \beta(i)$ is the positivity clause for the variable x_i, and (2) $\gamma = \alpha$ is a pairwise interaction clause linking (i, x_i) and (j, x_j).

Lemma 1. *For each positivity clause* $\beta(i)$*, we have* $R^*_{(i,x_i)\to\beta(i)} = 0$ *and* $M^*_{\beta(i)\to(i,x_i)} = 0$.

Proof. For positivity clauses, the set $C^u_{\beta(i)}(i, x_i)$ is the set of all the interaction clauses linked to (i, x_i), and the set $C^s_{\beta(i)}(i, x_i)$ is the empty set. (Refer to Figure 1(b) of the main paper for an illustration in a simple case). Among the interaction clauses, there is the single-variable clause $\gamma_{(i,x_i)}$ of the variable $\sigma_{(i,x_i)}$ (condition (3) in Definition 1), for which

$$\begin{aligned} M^s_{\gamma_{(i,x_i)}\to(i,x_i)} &= \exp(w_\gamma) \\ M^u_{\gamma_{(i,x_i)}\to(i,x_i)} &= \exp(-y_{\mathrm{src}(\gamma_{i,x_i})}) \\ M^*_{\gamma_{(i,x_i)}\to(i,x_i)} &= 0 \end{aligned}$$

From equation 9, since $C^s_{\beta(i)}(i, x_i) = \emptyset$ and $M^*_{\gamma_{(i,x_i)}\to(i,x_i)} = 0$, we have $R^*_{(i,x_i)\to\beta(i)} = 0$. From equation 6, this implies $M^*_{\beta(i)\to(i,x_i)} = 0$. ☐

On the other hand, for the interaction clauses, $M^*_{\alpha\to(i,x_i)} \neq 0$ in general. The update equations in the main paper are proved in the following theorem:

Theorem 1. *Let* $\beta(i)$ *be the positivity clauses of variables* (i, x_i)*, and* α *be the interaction clause linking* (i, x_i) *and* (j, x_j)*. Define*

$$\begin{aligned} \mu_{\beta(i)\to(i,x_i)} &= \frac{M^u_{\beta(i)\to(i,x_i)}}{M^s_{\beta(i)\to(i,x_i)}} \\ \nu_{\alpha\to(i,x_i)} &= \frac{M^u_{\alpha\to(i,x_i)}}{M^s_{\alpha\to(i,x_i)} + M^*_{\alpha\to(i,x_i)}} \end{aligned}$$

The update equations in Figure 1 can be rewritten using the messages μ *and* ν*.*

$$\mu_{\beta(i)\to(i,x_i)} = \sum_{x'_i\neq x_i} \prod_{\alpha\in N(i,x'_i)\setminus\beta(i)} \nu_{\alpha\to(i,x'_i)} + \exp(-w_i) \tag{13}$$

$$\nu_{\alpha\to(i,x_i)} = \frac{\mu_{\beta(j)\to(j,x_j)} + \exp(-y_{\mathrm{src}(\alpha)} - w_\alpha)\prod_{\gamma\in N(j,x_j)\setminus\{\beta(j),\alpha\}} \nu_{\gamma\to(j,x_j)}}{\mu_{\beta(j)\to(j,x_j)} + \prod_{\gamma\in N(j,x_j)\setminus\{\beta(j),\alpha\}} \nu_{\gamma\to(j,x_j)}} \tag{14}$$

$$B_{(i,x_i)}(0) \propto \mu_{\beta(i)\to(i,x_i)} \tag{15}$$

$$B_{(i,x_i)}(1) \propto \prod_{\alpha\in N(j,x_j)\setminus\beta(j)} \nu_{\alpha\to(j,x_j)} \tag{16}$$

$$B_{(i,x_i)}(*) = 0 \tag{17}$$

Proof. Let's work out the update equations for the interaction clauses first. For an interaction clause α linking variables (i, x_i) and (j, x_j), the set $C^s_\alpha(j, x_j)$ are all the other interaction clauses linked to (j, x_j), and the set $C^u_\alpha(j, x_j)$ contains the single positivity clause, $\beta(j)$. For a clause γ, denote $M^{s*}_{\gamma\to(j,x_j)} = M^s_{\gamma\to(j,x_j)} + M^*_{\gamma\to(j,x_j)}$, we have (from equations 7, 8, 9, and Lemma 1,

$$
\begin{aligned}
R^s_{(j,x_j)\to\alpha} &= M^u_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^{s*}_{\gamma\to(j,x_j)} \\
R^u_{(j,x_j)\to\alpha} &= M^s_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^u_{\gamma\to(j,x_j)} \\
R^*_{(j,x_j)\to\alpha} &= M^u_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^{s*}_{\gamma\to(j,x_j)}
\end{aligned}
$$

Following equations 4, 5 and 6, the messages from α to (i, v) can be written as follows

$$
\begin{aligned}
M^{s*}_{\alpha\to(i,x_i)} &= \exp(w_\alpha)\left(R^u_{(j,x_j)\to\alpha} + R^*_{(j,x_j)\to\alpha}\right) \\
&= \exp(w_\alpha)\left[M^s_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^u_{\gamma\to(j,x_j)} + M^u_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^{s*}_{\gamma\to(j,x_j)}\right] \\
M^u_{\alpha\to(i,x_i)} &= \exp(w_\alpha) R^s_{(j,x_j)\to\alpha} + \exp(-y_{\mathrm{src}(\alpha)}) R^u_{(j,x_j)\to\alpha} \\
&= \exp(w_\alpha)\left[M^u_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^{s*}_{\gamma\to(j,x_j)}\right] + \exp(-y_{\mathrm{src}(\alpha)})\left[M^s_{\beta(j)\to(j,x_j)} \prod_{\gamma\in C^s_\alpha(j,x_j)} M^u_{\gamma\to(j,x_j)}\right]
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\nu_{\alpha\to(i,x_i)} &= \frac{M^u_{\alpha\to(i,x_i)}}{M^{s*}_{\alpha\to(i,x_i)}} \\
&= \frac{\exp(w_\alpha)\mu_{\beta(j)\to(j,x_j)} + \exp(-y_{\mathrm{src}(\alpha)}) \prod_{\gamma\in C^s_\alpha(j,x_j)} \nu_{\gamma\to(j,x_j)}}{\exp(w_\alpha)(\prod_{\gamma\in C^s_\alpha(j,x_j)} \nu_{\gamma\to(j,x_j)} + \mu_{\beta(j)\to(j,x_j)})} \\
&= \frac{\mu_{\beta(j)\to(j,x_j)} + \exp(-y_{\mathrm{src}(\alpha)} - w_\alpha) \prod_{\gamma\in C^s_\alpha(j,x_j)} \nu_{\gamma\to(j,v')}}{\mu_{\beta(j)\to(j,x_j)} + \prod_{\gamma\in C^s_\alpha(j,x_j)} \nu_{\gamma\to(j,x_j)}}
\end{aligned}
$$

For the positivity clauses, the set $C^u_\alpha(j, x_j)$ are all the interaction clauses linked to (j, x_j), and the set $C^s_\alpha(j, x_j)$ is the empty set. By Lemma 1 and equations 7 and 8,

$$
\begin{aligned}
R^s_{(i,x_i)\to\beta(i)} &= \prod_{\gamma\in C^u_\alpha(i,x_i)} M^u_{\gamma\to(i,x_i)} \\
R^u_{(i,x_i)\to\beta(i)} &= \prod_{\gamma\in C^u_\alpha(i,x_i)} M^{s*}_{\gamma\to(i,x_i)} \\
R^*_{(i,v)\to\beta(i)} &= 0,
\end{aligned}
$$

From equations 4 and 5, taking into account the assumption that $y_i = 0$ for positivity clauses,

$$
\begin{aligned}
M^s_{\beta(i)\to(i,x_i)} &= \exp(w_i) \prod_{x_i'\neq x_i} R^u_{(i,x_i')\to\beta(i)} \\
M^u_{\beta(i)\to(i,x_i)} &= \exp(w_i) \sum_{x_i'\neq x_i}\left(R^s_{(i,x_i')\to\beta(i)} \prod_{x_i''\neq x_i', x_i} R^u_{(i,x_i'')\to\beta(i)}\right) + \prod_{x_i'\neq x_i} R^u_{(i,x_i')\to\beta(i)} \\
\mu_{\beta(i)\to(i,x_i)} &= \sum_{x_i'\neq x_i} \frac{R^s_{(i,x_i')\to\beta(i)}}{R^u_{(i,x_i')\to\beta(i)}} + \exp(-w_i) \\
&= \sum_{x_i'\neq x_i} \prod_{\gamma\in C(i,x_i')-\{\beta(i)\}} \nu_{\gamma\to(i,x_i')} + \exp(-w_i)
\end{aligned}
$$

To express the beliefs in terms of the μ and ν messages, we apply equations 10 to 12 for the case where $C^+(i)$ is the singleton $\beta(i)$, and $C^-(i)$ is the set of all interaction clauses linked to the variable i. It is easy to see that in this case, the equations 10 to 12 are proportional to the equations 15 to 17. $\square$

We found empirically that the asynchronous schedule of message updates affect convergence to a large extent. A good schedule for message updates is to update all the ν-messages first (by updating the groups of ν-messages belonging to each factor $a \in F$ together), and then updating the μ-messages together. This schedule seems to work better than the schedule defined by residual belief propagation [1] on the relaxed MRF.

References

[1] G. Elidan, I. McGraw, and D. Koller, "Residual belief propagation: Informed scheduling for asynchronous message passing," in *UAI*, 2006.

[2] E. Maneva, E. Mossel, and M. Wainwright, "A new look at survey propagation and its generalizations," 2004. [Online]. Available: http://arxiv.org/abs/cs.CC/0409012