Submitted for consideration for ISIT 2003, October 30, 2002.

Poset Belief Propagation–Experimental Results*

Jonathan Harel, Robert J. McEliece, and Ravi Palanki
California Institute of Technology
Pasadena, California USA

Abstract: Poset belief propagation, or PBP, is a flexible generalization of ordinary belief propagation which can be used to design algorithms for solving (approximately) many probabilistic inference problems, including MAP decoding of binary linear codes. In this paper, we will present some experimental results tha suggest that PBP can significantly outperform conventional BP techniques.

1. Introduction.

In [4], building on the pioneering work of Yedidia, Freeman, and Weiss [10] on "generalized belief propagation," McEliece and Yildirim introduced a class of algorithms called *belief propagation on partially ordered sets*, or PBP. (Similar algorithms have recently been developed in [11] and [5, 6]). PBP includes as special cases ordinary belief propagation [7], probability propagation [8], the generalized distributive law [1, 2], the sum-product algorithm [3], generalized belief propagation [10], (all of these with and without loops), and many other instances whose effectiveness has not yet been investigated in detail. In this paper we summarize PBP (including a new "message-free" formulation of the algorithm) and report the results of some experiments we have performed. Our tentative conclusion is that PBP can often outperform ordinary BP. This suggests that PBP may prove to be effective when ordinary BP is not. By the time of ISIT 2003, we expect to be able to report the results of numerous additional experiments, including experiments with PBP-based decoding algorithms.

2. The Marginalized Product Density Problem.

Technically, PBP is an algorithm for solving any *marginalized product density* problem. In this section we describe the MPD problem.

Let $\{x_1, \ldots, x_n\}$ be a collection of n variables taking values in the finite set $A = \{0, 1, \ldots, q-1\}$, and Let $\mathcal{R} = \{R_1, \ldots, R_M\}$, be M sparse subsets of $\{1, 2, \ldots, n\}$. Suppose we are given a set of nonnegative "local potentials" $\{\alpha_R(\boldsymbol{x}_R) : R \in \mathcal{R}\}$. (Here we have introduced the notation $\boldsymbol{x}_R$ to denote the components of $\boldsymbol{x}$ which have indices in R).

These kernels define, in principle, a product probability density function;

$$B(\boldsymbol{x}) = \frac{1}{Z} \prod_{R \in \mathcal{R}} \alpha_R(\boldsymbol{x}_R),$$

where Z is chosen so that $\sum_{\boldsymbol{x} \in A^n} B(\boldsymbol{x}) = 1$.

* This research was supported by NSF grant no. CCR-0118670, and grants from Sony, Qualcomm, and Caltech's Lee Center for Advanced Networking.

The marginalized product density problem is to compute, exactly or approximately, some or all of the *local marginal densities* $\{B_R(\boldsymbol{x}_R)\}_{R\in\mathcal{R}}$ of the product density, where

$$B_R(\boldsymbol{x}_R) = \sum_{\boldsymbol{x}\setminus\boldsymbol{x}_R} B(\boldsymbol{x}).$$

Appropriately interpreted, the MPD problem includes as special cases turbo/LDPC decoding, probabilistic inference in Bayesian networks, certain problems in machine vision, finite Fourier transforms, and free energy computations in statistical physics.

3. Posets and Junction Posets.

Let P be a finite partially ordered set (see [4,9] for details). The *Hasse diagram* $H = H(P)$ for P is a graph with vertex set P, with (ρ, σ) being an edge in H if and only if $\rho > \sigma$ and if there is no element τ such that $\rho > \tau > \sigma$. We represent the ordering of an edge $e = (\rho, \sigma)$ by placing ρ "above" σ in the Hasse diagram. We say that the poset P is *connected* if the Hasse diagram H is connected. Figures 1-3 show the Hasse diagrams of three small connected posets.

We assign an *overcounting number* $\phi(\rho)$ to each $\rho \in P$, such that

$$\sum_{\rho:\rho\geq\sigma} \phi(\rho) = 1, \text{ for all } \sigma \in P. \tag{3.1}$$

The numbers $\phi(\rho)$ are integers and are determined uniquely by (3.1). In Figs. 1–3, the overcounting numbers are shown as labels on the vertices of the Hasse diagrams.

Given a collecton $\mathcal{R}$ of subsets of $\{1, \ldots, n\}$ as in Section 2, a poset P is called a *junction poset for* $\mathcal{R}$ if each element $\rho \in P$ is assigned a subset $L(\rho) \subseteq \{1, \ldots, n\}$ and a subset $\mathcal{R}(\rho) \subseteq \mathcal{R}$, such that:

$$\bigcup_{R\in\mathcal{R}(\rho)} R \subseteq L(\rho) \qquad \text{for all } \rho \in P$$

$$\mathcal{R}(\sigma) \subseteq \mathcal{R}(\rho) \qquad \text{if } \sigma \leq \rho,$$

and such that the following conditions are satisfied:

$$P_i = \{p \in P : i \in L(\rho)\} \quad \text{is a connected poset, for all } i = 1, \ldots, n. \tag{3.2}$$

$$\sum_{\rho: i\in L(\rho)} \phi(\rho) = 1, \text{ for all } i \in \{1, \ldots, n\} \qquad (\text{"1-variable balance"}) \tag{3.3}$$

$$\sum_{\rho: R\in\mathcal{R}(\rho)} \phi(\rho) = 1, \text{ for all } R \in \mathcal{R}. \qquad (\text{"conservation of energy"}) \tag{3.4}$$

In Figs. 1–3 we have labelled each vertex with a subset of $\{1, 2, 3, 4, 5\}$.

4. The Poset-BP Algorithm.

We assume that P is a junction poset for $\mathcal{R}$. Throughout the algorithm, each vertex $\rho \in P$ carries a "belief table" (probability density) $b_\rho(\boldsymbol{x}_\rho)$, where $\boldsymbol{x}_\rho$ is shorthand for $\boldsymbol{x}_{L(\rho)}$. Initially, the belief at ρ is defined as follows:

$$b_\rho(\boldsymbol{x}_\rho) = \prod_{R\in\mathcal{R}(\rho)} \alpha_R(\boldsymbol{x}_R).$$

The beliefs are then updated by keeping the product function

$$Q(\boldsymbol{x}) = \prod_{\rho \in P} b_\rho(\boldsymbol{x}_\rho)^{\phi(\rho)},$$

which is initially proportional to $B(\boldsymbol{x})$, invariant, while *enforcing edge consistency.*

An edge $e = (\rho, \sigma)$ in the Hasse diagram is said to be *consistent* if

$$\sum_{\boldsymbol{x}_\rho \setminus \boldsymbol{x}_\sigma} b_\rho(\boldsymbol{x}_\rho) = b_\sigma(\boldsymbol{x}_\sigma) \qquad \text{for all } \boldsymbol{x}_\sigma \in A^{L(\sigma)}.$$

The iteration step of the PBP algorithm is as follows. If the edge $e = (\rho, \sigma)$ is inconsistent, define the "correction table" as

$$\Delta_e(\boldsymbol{x}_\sigma) = \frac{\sum_{\boldsymbol{x}_\rho \setminus \boldsymbol{x}_\sigma} b_\rho(\boldsymbol{x}_\rho)}{b_\sigma(\boldsymbol{x}_\sigma)},$$

and update the beliefs as follows:

$$b_\tau(\boldsymbol{x}_\tau) \leftarrow b_\tau(\boldsymbol{x}_\tau) \cdot \Delta_e(\boldsymbol{x}_\sigma), \tag{4.1}$$

for all τ such that $\tau \geq \sigma$, $\tau \not\geq \rho$. (But see (5.1), below.) This step makes e consistent (take $\tau = \sigma$), and preserves $Q(\boldsymbol{x})$ (since $\sum_{\tau : \tau \geq \sigma, \tau \not\geq \rho} \phi(\tau) = 0$).

Beliefs are updated until all edges are consistent, at which point the algorithm has reached a fixed point and the beliefs $\{b_\rho(\boldsymbol{x}_\rho)\}_{\rho \in P}$ are (hopefully) close to the desired marginals, i.e.,

$$b_\rho(\boldsymbol{x}_\rho) \approx B_\rho(\boldsymbol{x}_\rho) = \sum_{\boldsymbol{x} \setminus \boldsymbol{x}_\rho} B(\boldsymbol{x}).$$

In the next section we report the results of some experiments we conducted to test this hope.

5. Experimental Results.

To test the effectiveness of PBP, we conducted a series of experiments, with $n = 5$ and $A = \{0, 1\}$. The local domains were

$$\mathcal{R} = \{\{1,2,3\}, \{1,3,4\}, \{2,3,5\}, \{3,4,5\}\},$$

with corresponding local potentials

$$\alpha_1(x_1, x_2, x_3),\ \alpha_2(x_1, x_3, x_4),\ \alpha_3(x_2, x_3, x_5),\ \alpha_4(x_3, x_4, x_5)$$

There are many possible junction posets for $\mathcal{R}$; we investigated the three shown in Figures 1–3. We ran PBP on each of the three posets for a thousand random choices of the local potentials, each time computing the five pairs (b_i, b_i^*), the exact and PBP approximations to the "belief" $B(x_i)$. The results (for $i = 1$) are shown in Figs. 4–6. (The tighter the clusterinng around the line $b_1 = b_1^*$, the better the performance.)

For poset 3, we found that PBP failed to converge with the update rule (4.1), but when we used either of the modified rules

$$
\begin{aligned}
b_\tau(\boldsymbol{x}_\tau) &\leftarrow b_\tau(\boldsymbol{x}_\tau) \cdot \Delta_e(\boldsymbol{x}_\sigma)^w \\
b_\tau(\boldsymbol{x}_\tau) &\leftarrow b_\tau(\boldsymbol{x}_\tau)((1-w) + w\Delta_e(\boldsymbol{x}_\sigma)),
\end{aligned}
\tag{5.1}
$$

wtih $w < 1$, the excellent performance shown in Figure 6 resulted.

6. Conclusions.

Based on the experiments reported here (and on similar experiments not reported) we conclude that for a given complexity, PBP on a carefully chosen poset is likely to outperform ordinary BP. (Notice, incidentally, that a comparison of Figures 4 and 5 indicates a significant difference between the GDL and the SPA. This is worth investigating.) This suggests that PBP may prove to be effective when ordinary BP is not. By the time of ISIT 2003, we expect to be able to report the results of a number of additional experiments, including experiments involving PBP decoding of some nontrivial linear codes.

References.

1. S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, no. 2 (March 2000), pp. 325–343.
2. S. M. Aji and R. J. McEliece, "The generalized distributive law and free energy minimization," Proc. 2001 Allerton Conf. Comm. Control and Computing (Oct. 2001).
3. F. R. Kschichang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47 no. 2 (Feb. 2001), pp. 498–519.
4. R. J. McEliece and M. Yildirim, "Belief Propagation on Partially Ordered Sets," to appear in the IMA Volume "Mathematical Systems Theory in Biology, Communication, Computation, and Finance," David Gilliam and Joachim Rosenthal, eds.
5. P. Pakzad and V. Anantharam, "Belief propagation and statistical physics," Proc. 2002 Conf. Inform. Sciences and Systems, Princeton U. March 2002.
6. P. Pakzad and V. Anantharam, "Minimal graphical representation of Kikuchi regions," preprint.
7. J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kaufmann, 1988.
8. G. R. Shafer and P. P. Shenoy, "Probability propagation," *Ann. Mat. Art. Intell.*, vol. 2 (1990), pp. 327–352.
9. R. P. Stanley, *Enumerative Combinatorics, vol. I.* (Cambridge Studies in Advanced Mathematics 49) Cambridge: Cambridge University Press, 1997.
10. J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," pp. 689–695 in *Advances in Neural Information Processing Systems 13*, (2000) eds. Todd K. Leen, Thomas G. Dietterich, and Volker Tresp.
11. J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," submitted to IEEE Trans. Inform. Theory, Sept. 2002. available at `www.merl.com/papers/ TR2002-xx/`

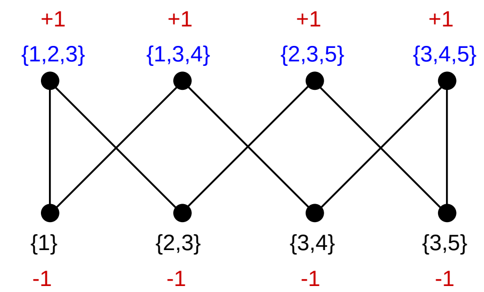

Figure 1. Poset 1 (junction graph construction [2]),
showing overcounting numbers and local domains.
Local potentials are assigned to the maximal elements.

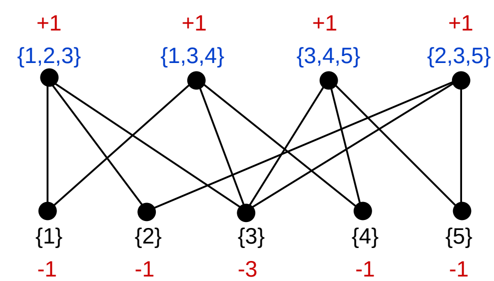

Figure 2. Poset 2 (factor graph construction [3]),
showing overcounting numbers and local domains.
Local potentials are assigned to the maximal elements.

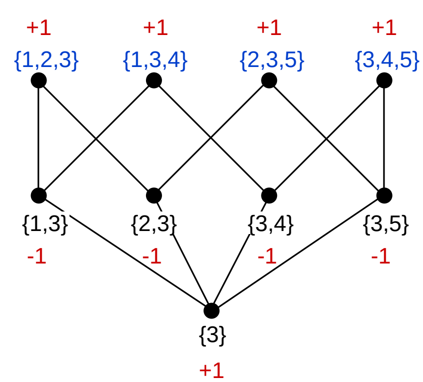

Figure 3. Poset 3 (cluster variational method [11]),
showing overcounting numbers and local domains.
Local potentials are assigned to the maximal elements.

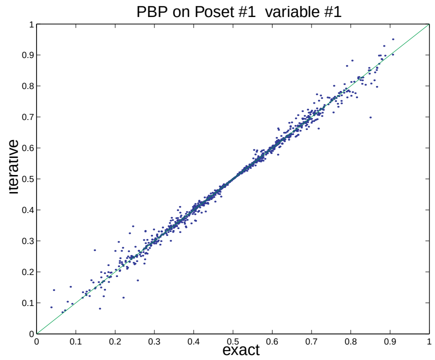

Figure 4. Experimental Results — Poset 1.

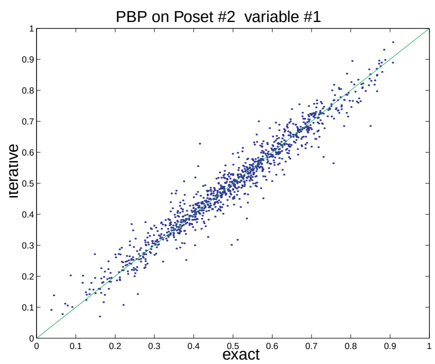

Figure 5. Experimental Results — Poset 2.

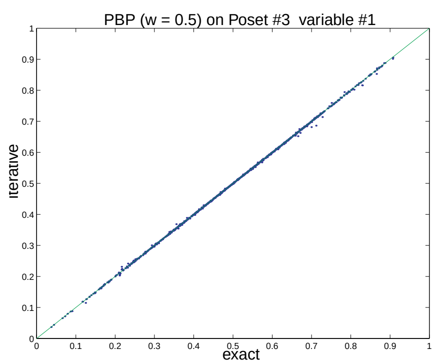

Figure 6. Experimental Results — Poset 3.