A comparison of Gibbs, Bethe, and Kikuchi free energies using Monte Carlo simulations

Delbert D. DUECK †, Brendan J. FREY †

***Abstract*— The sum-product algorithm is an important tool for performing probabilistic inference. It has been shown that when applied to cyclic factor graphs (loopy belief propagation), the stationary points of the algorithm correspond to minima of the network's Bethe free energy, a concept from statistical physics. The Kikuchi free energy, a later generalization of the Bethe free energy, is believed to better approximate the Gibbs free energy (a common cost function) than the Bethe free energy. This paper calculates and compares all three quantities for trivial-sized Ising spin-glass factor graphs sampled Monte Carlo-style from the probability space, and finds that the Kikuchi approximation is better roughly 60-80% of the time.**

I. Introduction

TECHNIQUES of approximate probabilistic inference are today being applied to a wide variety of problems with large and/or high-dimensional data sets. In the context of these, graphical models such as the factor graph are used to convey ideas, as opposed to lengthy textual or mathematical descriptions.

A common one encounters with graphical models containing networks of random variables is calculating marginal probability distributions for individual nodes. A traditional statistical approach is to sum over all variables except for the variable of interest, however this becomes intractable beyond 10 or 20 variables. The sum product algorithm [1] overcomes this limitation by strategically shifting independent factors outside of these summations in the following manner:

$$\begin{aligned}&\sum_x (\text{factors indep. of } x)(\text{factors dep. on } x) = \\ &\quad (\text{factors indep. of } x)\sum_x (\text{factors dep. on } x)\end{aligned} \tag{1}$$

For acyclic networks, efficient use of this identity can perform an exponential-time complete marginalization problem in linear time. Equally dramatic running-time improvements are also possible in cyclic graphs (loopy belief propagation), however, the user must settle for approximate results.

Through most of the 1990's, the only justification behind this loopy belief propagation was mostly-positive experimental results. This changed when Jonathan Yedidia [2] – borrowing decades-old ideas from theoretical physics – showed that applying the sum-product algorithm to arbitrary graph structures was equivalent to minimizing H.A. Bethe's 1935 approximation to the free energy.

Manuscript received October 2, 2003. This work was supported in part by the Probabilistic and Statistical Inference research laboratory at the University of Toronto.

† Electrical and Computer Engineering, University of Toronto, 10 King's College Rd., Toronto, Ontario, Canada M5S 3G4. www.psi.toronto.edu

II. Free Energies and their Approximations

A. Gibbs Free Energy

A common measurement of "distance" between a true probability distribution $P(x)$ and an approximating distribution $Q(x)$ is the Kullback-Leibler (KL) divergence, defined as follows:

$$D\big(Q(x)\,\|\,P(x)\big) = \sum_x Q(x)\cdot\log\frac{Q(x)}{P(x)} \tag{2}$$

Yedidia shows in [3] that the KL-divergence has close parallels with the Gibbs free energy from Bolzmann's law in theoretical physics. The exact joint distribution $P(\mathbf{x})$ for a factor graph defined by potential functions $\phi_i(x_{C_i})$ over clusters C_i is:

$$P(\mathbf{x}) \propto \Phi(\mathbf{x}) = \prod_i \phi_i(x_{C_i}) \tag{3}$$

The Gibbs free energy, obtained by substituting Bolzmann's law into the KL-divergence definition, is:

$$\sum_{\mathbf{x}} Q(\mathbf{x})\cdot\log\frac{Q(\mathbf{x})}{\Phi(\mathbf{x})} \tag{4}$$

For the special case of $P(\mathbf{x}) = \Phi(\mathbf{x})$, where the factor graph potentials are pre-normalized to sum to one, the Gibbs free energy is equal to the KL-divergence. Otherwise, it exceeds the KL-divergence by a constant, known as the Helmholz free energy.

B. Free Energy Approximations

Practically, inference problems often involve approximately marginalizing $P(\mathbf{x})$ to the believed distribution $Q(\mathbf{x})$. For random variable networks of non-trivial size, finite memory prevents the tabulation of joint distributions and, consequently, the Gibbs free energy.

The Bethe free energy [4] is a useful approximation of the Gibbs free energy that only requires marginal distributions and cluster potentials (often pair-wise distributions in practice). Yedidia shows in [3] that minimizing the Bethe free energy is equivalent to performing the sum-product algorithm; stationary points to which the sum-product algorithm converges correspond to local minima of the Bethe free energy. It is calculated as follows:

$$F_{Bethe} = \sum_i \sum_{x_{C_i}} q_{C_i}(x_{C_i}) \cdot \left(\log q_{C_i}(x_{C_i}) + \log \phi_{C_i}(x_{C_i})\right) - \sum_i (d_i - 1) \sum_{x_i} q_i(x_i) \cdot \left(\log q_i(x_i) + \log \phi_i(x_i)\right) \quad (5)$$

In this case, the approximate distribution $Q(\mathbf{x})$ mirrors the structure of $P(\mathbf{x})$ from (3); q_{C_i} are the cluster marginal probabilities and q_{C_i} are the individual variable node marginals. The d_i counting numbers refer to the number of neighbors each variable node has – its use parallels Möbius numbers described in [5], preventing variable nodes from being over-counted.

The Bethe free energy approximation was generalized in Ryoichi Kikuchi's 1951 paper [6] to a higher-order approximation, making use of multi-level clustering. His paper leaves Bethe's approximation as the second-order special case (containing one cluster level and the individual node level) to his algorithm. For the simulations described in this paper, third-order Kikuchi approximations are used.

C. *Ising Spin-Glass model*

The factor graph of interest in this paper is known as the Ising spin-glass model [7]. It is commonly used to model ferromagnetic domain alignments. A sample 5×5 network is shown in Figure 1.

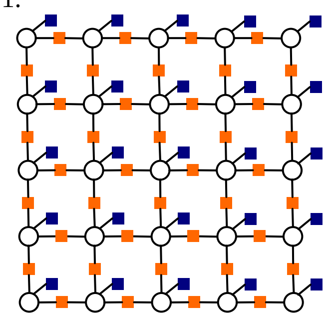

Figure 1: Ising Spin-Glass Factor Graph

The network follows usual factor graph convention, with filled squares representing function nodes and unfilled circles representing variable nodes. With the Bethe approximation, the individual node free energies are multiplied by $d_i - 1$ and subtracted to prevent over-counting (due to pair-wise clusters) as follows:

$$\begin{matrix} 1 & 2 & 2 & \cdots & 2 & 2 & 1 \\ 2 & 3 & 3 & 3 & 3 & 3 & 2 \\ 2 & 3 & 3 & 3 & 3 & 3 & 2 \\ \vdots & 3 & 3 & \cdots & 3 & 3 & \vdots \\ 2 & 3 & 3 & 3 & 3 & 3 & 2 \\ 2 & 3 & 3 & 3 & 3 & 3 & 2 \\ 1 & 2 & 2 & \cdots & 2 & 2 & 1 \end{matrix}$$

Figure 2: Bethe individual counting numbers d_i–1

The Kikuchi approximation is third-order, using (square) quaternary clustering, pair-wise clustering, and individual nodes. This scheme is illustrated in Figure 3.

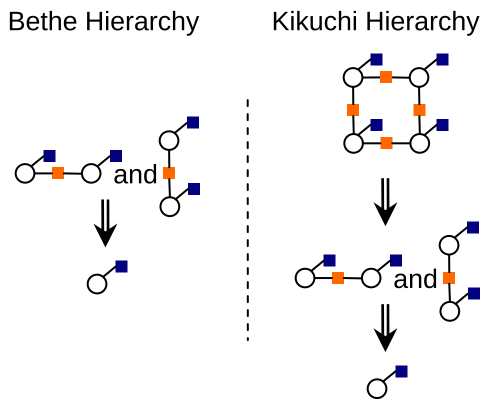

Figure 3: Bethe and Kikuchi hierarchies

The pair-wise Kikuchi counting numbers are as follows:

$$\begin{matrix} 0 & 0 & \cdots & 0 & 0 \\ -1 & -1 & \cdots & -1 & -1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ -1 & -1 & \cdots & -1 & -1 \\ 0 & 0 & \cdots & 0 & 0 \end{matrix} \qquad \begin{matrix} 0 & -1 & \cdots & -1 & 0 \\ 0 & -1 & \cdots & -1 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & -1 & \cdots & -1 & 0 \\ 0 & -1 & \cdots & -1 & 0 \end{matrix}$$

Figure 4: Kikuchi pair-wise counting numbers – horizontal pair-wise (left) and vertical pair-wise (right)

Note that for an Ising grid of size M×N, the horizontal and vertical pair-wise counting numbers form M×(N–1) and (M-1)×N matrices, respectively. The Kikuchi individual variable-node counting numbers, forming an N×M matrix, are shown in Figure 5.

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \cdots & 1 & 0 \\ 0 & \vdots & \cdots & \vdots & 0 \\ 0 & 1 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Figure 5: Kikuchi individual node counting numbers

III. Simulation Results

Binary-valued Ising grids of sizes 3×3, 4×4, and 5×5 were constructed in MATLAB by randomly initializing the singular and horizontal / vertical pair-wise potentials, and multiplying the terms as in (3) to form a joint distribution.

$$P(\mathbf{x}) \propto \Phi(\mathbf{x}) = \prod_{\substack{1\le m\le M,\\ 1\le n\le N}} \phi_{mn} \prod_{\substack{1\le m\le M,\\ 1\le n\le N-1}} \phi_{mn}^{\leftrightarrow} \prod_{\substack{1\le m\le M-1,\\ 1\le n\le N}} \phi_{mn}^{\updownarrow} \quad (6)$$

This "true" joint distribution was marginalized to determine the true individual marginals $p(x_{mn})$, the true pair-wise distributions $p(x_{mn}^{\leftrightarrow})$ and $p(x_{mn}^{\updownarrow})$, and the true quaternary distributions $p(x_{mn}^{\square})$.

For the simulations, many approximate joint distributions $Q(\mathbf{x})$ were formed by Monte Carlo sampling of the probability space in the rough area of $P(\mathbf{x})$. The sampling was constructed so they deviated from the true distribution analogous to multiplicative Gaussian noise:

$$Q(\mathbf{x}) = P(\mathbf{x})n(\mathbf{x}), \quad n(x) \sim \mathcal{N}(1, \sigma^2) \tag{7}$$

The new $Q(\mathbf{x})$ was then marginalized in a manner similar to $P(\mathbf{x})$. The Gibbs, Bethe, and Kikuchi free energies were calculated using the $Q(\mathbf{x})$, $q(x_{mn})$, $q(x_{mn}^{\leftrightarrow})$, $q(x_{mn}^{\updownarrow})$, and $q(x_{mn}^{\square})$ terms. It was then tabulated whether the Bethe free energy or the Kikuchi free energy was a better approximation to the Gibbs free energy – their absolute errors are listed as $|\varepsilon_{Bethe}|$ and $|\varepsilon_{Kikuchi}|$. The results are as follows:

Table 1: Simulation Results for Consistent Beliefs

σ	Grid size	Count of $\|\varepsilon_{Bethe}\|<\|\varepsilon_{Kikuchi}\|$	Count of $\|\varepsilon_{Kikuchi}\|<\|\varepsilon_{Bethe}\|$	Relative freq. of $\|\varepsilon_{Bethe}\|<\|\varepsilon_{Kikuchi}\|$
.01	3×3	38	962	3.8%
.01	4×4	1	99	1%
.01	5×5	0	100	0%
.10	3×3	212	788	21.2%
.10	4×4	34	66	34%
.10	5×5	40	60	40%
.25	3×3	188	812	18.8%
.25	4×4	48	52	48%
.25	5×5	61	39	61%
1.00	3×3	63	937	6.3%
1.00	4×4	21	79	21%
1.00	5×5	51	49	51%

It is surprising how often the Bethe free energy better-approximates the Gibbs free energy – the error measure – compared to the presumably-superior Kikuchi free energy. With large multiplicative noise variances and a 5×5 Ising grid, some configurations even show the Bethe free energy prevailing most of the time – this is probably due to the small number of trials (100, due to time constraints) rather than a meaningful result.

A scatter plot of the 100 samples for the 4×4 Ising grid results (at four different multiplicative noise powers, σ^2) is shown in Figure 6.

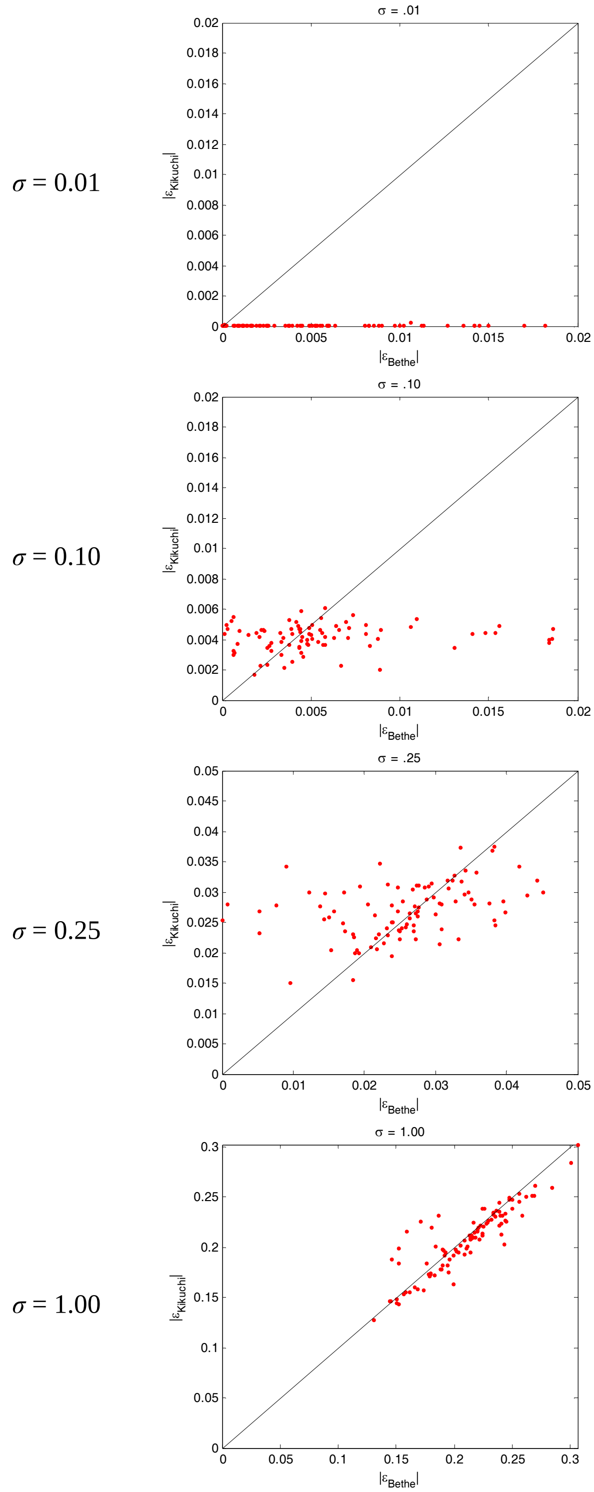

Figure 6: $|\varepsilon_{Kikuchi}|$ vs. $|\varepsilon_{Bethe}|$ for 4x4 grids

In the previous plots, the absolute Kikuchi free energy error ($|\varepsilon_{Kikuchi}|$) is plotted against the absolute Bethe free energy error ($|\varepsilon_{Bethe}|$). The surprising trial results – those where Bethe's approximation outperforms Kikuchi's approximation – are the data points located above the $y = x$ line.

IV. Discussion

This paper presented some preliminary results of Monte

Carlo simulations comparing the Bethe and Kikuchi free energy approximations to the true Gibbs free energy that merit further investigation.

The Bethe approximation is often viewed as a second-order special case of the generalized n^{th}-order Kikuchi approximation. An analogy can be made to the infinite series expansion of polynomials; this is much in the same way that a second-order MacLaurin series polynomial expansion (i.e. $f(x) = f(0) + xf'(0) + \frac{1}{2}x^2 f''(0)$) is seen as a second-order special case of the general series ($\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0)x^k$).

The analogy stops when the convergence patterns of the two are compared. MacLaurin series expansions of well-behaved functions always do better as more terms are included. This paper shows that increasing the complexity of the Kikuchi approximation – or at least the transition from 2^{nd}-order to 3^{rd}-order – is in no way a guarantee of increased accuracy. The results in this paper suffer from small Monte Carlo sample size (e.g. 100 samples), so no hard conclusions can be drawn as to how often or in what situations the Bethe approximation tends to outdo the 3^{rd}-order Kikuchi approximation – the only firm conclusion is that the rate is not astronomically small.

This raises some concerns about the benefit of generalized belief propagation algorithms (GBP), which minimize the Kikuchi free energy. They purportedly have better performance than the standard sum-product loopy belief propagation algorithm (which minimizes the Bethe free energy). Superior performance is to be expected if the error measure for this judgment is the Kikuchi free energy, but the Gibbs free energy angle should be examined too.

This is in addition to the well-document “overshooting” problem in GBP [3]. Message update inertia must be significantly lowered by giving heavy weighting to the previous (“old) message values in update equations, lest convergence problems materialize.

Further work comparing Bethe and Kikuchi approximations – in different models of larger size – is warranted.

References

[1] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor Graphs and the sum-product algorithm. *IEEE Trans. Info. Theory*, 47:498-519, 2001.

[2] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13, Cambridge, MA, 2001. MIT Press.

[3] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms. Available at http://www.merl.com/reports/TR2002-35/, 2002.

[4] H. A. Bethe. *Proc. Roy. Soc. London A*, 150:552, 1935.

[5] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding Belief Propagation and its Generalizations. Available at http://www.merl.com/reports/TR2001-22/, 2001.

[6] R. Kikuchi. A theory of cooperative phenomena. *Phys. Rev.*, 81(6):988-1003, 1951.

[7] M. Mezard, G. Parisi, and M. A. Virasoro. Spin glass theory and neyond. *World Scientific*, 1987.