Fixed Point Solutions of Belief Propagation

Christian Knoll
Signal Processing and Speech Comm. Lab.
Graz University of Technology
christian.knoll@tugraz.at

Franz Pernkopf
Signal Processing and Speech Comm. Lab.
Graz University of Technology
pernkopf@tugraz.at

Dhagash Mehta
Dept of ACMS
University of Notre Dame, USA.
dmehta@nd.edu

Tianran Chen
Dept of Mathematics and Computer Science
Auburn University at Montgomery
ti@nranchen.org

Abstract

Belief propagation (BP) is an iterative method to perform *approximate* inference on arbitrary graphical models. Whether BP converges and if the solution is a unique fixed point depends on both, the structure and the parametrization of the model. To understand this dependence it is interesting to find *all* fixed points. In this work, we formulate a set of polynomial equations, the solutions of which correspond to BP fixed points.

To solve such a nonlinear system we present the numerical polynomial-homotopy-continuation (NPHC) method. We apply the proposed method to obtain all BP fixed points on binary Ising models. Further we compare the accuracy of the corresponding marginals to the exact marginal distribution. Contrary to the conjecture that uniqueness of BP fixed points implies convergence, we find graphs for which BP fails to converge, even though a unique fixed point exists. Moreover, we show that this fixed point gives a good approximation, and the NPHC method is able to obtain this fixed point.

1 Introduction

Performing exact inference on general probabilistic graphical models with many random variables (RVs) is computational intractable. Belief propagation (BP) is an efficient method to approximate the marginal distribution on graphs with loops. Despite the lack of guarantee for convergence, BP has been successfully used for models with many loops, including applications in computer vision, medical diagnosis systems, and speech processing [18, 1, 19]. However, instances of graphs do exist where BP fails to converge. A deeper understanding of the reasons for convergence of BP, and whether and how its non-convergence relates to the number of fixed points may therefore be crucial in understanding BP.

Results that relate accuracy and convergence rate are available for small grid graphs and graphs with a single loop [8, 22]. Sufficient conditions for uniqueness of fixed points were refined by accounting for both, the potentials as well as the structure of the model [7, 17]. Still the precise relation among the uniqueness of fixed points, convergence rate, and accuracy is yet to be theoretically understood [17].

In this work we aim to find *all* fixed points of BP to get deeper insights into the behavior of BP. If BP converges it does only provide a *single* fixed point, whereas the recently proposed approximation of survey propagation [21] represents distributions over BP messages. However, in order to find *all* fixed points – including unstable ones, we reformulate the update rules of BP as a system of polynomial equations.

There are indeed several methods to solve such a system of polynomial equations. Numerical solvers (e.g., Newton's method) are well established, but their ability in obtaining the solutions strongly depends on the initial point. Moreover, such methods only find a single solution at a time. Symbolic methods, on the other hand, are guaranteed to find *all* solutions. The Gröbner basis method [4, 5] is widely used, but it is inefficient if the system has irrational coefficients. Furthermore it suffers from fast growing run time and memory complexity, and has a limited scalability in parallel computation. In this paper, we present the numerical polynomial homotopy continuation (NPHC) method [11, 20] that overcomes all the above mentioned problems of both iterative and symbolic methods, yet guarantees to find *all* solutions of the system. Over the last few decades, this class of methods has been proven to be robust, efficient, and highly parallelizeable.

We apply the NPHC method to different realizations of the Ising model. On these graphs we obtain all BP fixed points and show how the number of fixed points changes at critical regions (i.e., phase transitions) in the parameter space. We further use the obtained fixed points to estimate the approximation of the marginal distribution which we compare to the exact marginal distribution. Our main (empirical) observations are: (i) on loopy graphs BP does not necessarily converge to the best possible fixed point, (ii) on some graphs where BP does not converge we are able to show that there is a unique fixed point, and (iii) if we enforce convergence to this unique fixed point the obtained marginals still give a good approximation.

2 Belief Propagation (BP)

We consider a finite set of N discrete random variables $\mathbf{X} = \{X_1, \dots, X_N\}$with the corresponding undirected graph $G = (\mathbf{X}, \mathbf{E})$, where $\mathbf{X} = \{X_1, \dots, X_N\}$ is the set of nodes and $\mathbf{E}$ is the set of edges. The joint distribution takes the form of the product $P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z}\prod_{l=1}^{L} \Phi_{C_l}(C_l)$, where the potentials Φ_{C_l} are specified over the maximal cliques C_l and $Z \in \mathbb{R}_+^*$ is a positive normalization constant [18, p.105]. When applying BP all potentials are restricted to consist of at most two variables: then, the joint distribution is factorized according to the normalized product

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_{(i,j)\,:\,e_{i,j}\in\mathbf{E}} \Phi_{X_i,X_j}(x_i, x_j) \prod_{i=1}^{N} \Phi_{X_i}(x_i). \tag{1}$$

The first product runs over all edges and the second product runs over all nodes; pairwise potentials and local evidence are denoted as Φ_{X_i,X_j} and Φ_{X_i} respectively. The marginals are approximated by recursively updating messages between all RVs. At iteration $n + 1$ the message between X_i and X_j for state x_j is given according to the update rule:

$$\mu_{i,j}^{n+1}(x_j) = \alpha_{i,j}^{n} \sum_{x_i \in \mathbb{S}} \Phi_{X_i,X_j}(x_i, x_j) \Phi_{X_i}(x_i) \prod_{X_k \in \Gamma_{i,j}} \mu_{k,i}^{n}(x_i), \tag{2}$$

where $\Gamma_{i,j} = \{X_k \in \mathbf{X} \backslash \{X_i, X_j\} : e_{i,j} \in \mathbf{E}\}$. Loosely speaking BP collects all messages sent to X_i, except from X_j, and multiplies this product with the local potential $\Phi_{X_i}(x_i)$ and the pairwise potential $\Phi_{X_i,X_j}(x_i, x_j)$. The sum over both states of X_i is sent to X_j. In practice the messages are often normalized by $\alpha_{i,j}^{n} \in \mathbb{R}_+^*$ so as to sum to one [9].

The set of all messages at iteration n is given by $\underline{\mu}^n = \{\mu_{i,j}^n(x_j) : e_{i,j} \in \mathbf{E}\}$. In a similar manner we collect all normalization terms in $\underline{\alpha}^n$. Let the mapping of all messages induced by (2) be denoted as $\underline{\mu}^{n+1} = BP\{\underline{\mu}^n\}$. If all successive messages show the same value (up to some predefined precision), then BP is converged and returns the approximated marginals $\tilde{P}(X_i = x_i) = \frac{1}{Z_i}\Phi_{X_i}(x_i)\prod_{X_k\in\partial(X_i)} \mu_{k,i}^*(x_i)$, where X_k are the neighbors of X_i, i.e., $\partial(X_i)$. We refer to converged messages and the associated normalization terms as *fixed points* $(\underline{\mu}^*, \underline{\alpha}^*)$.

3 Solving BP Fixed Point Equations

To find all fixed points of BP we are naturally interested in conditions such that $BP\{\underline{\mu}^n\} = \underline{\mu}^n$. Considering binary RVs, we estimate the residual, i.e., the difference between two successive message values. Therefore, we formulate the following system of polynomial equations, whose solutions are the fixed points of BP.

$$
\mathbf{F}(\underline{\mu},\underline{\alpha}) = \begin{cases} -\mu^n_{i,j}(x_j) + \alpha^n_{i,j} \sum\limits_{x_i \in \mathbb{S}} \Phi_{X_i,X_j}(x_i,x_j)\Phi_{X_i}(x_i) \prod\limits_{X_k \in \Gamma_{i,j}} \mu^n_{k,i}(x_i) \\ -\mu^n_{i,j}(\bar{x}_j) + \alpha^n_{i,j} \sum\limits_{x_i \in \mathbb{S}} \Phi_{X_i,X_j}(x_i,\bar{x}_j)\Phi_{X_i}(x_i) \prod\limits_{X_k \in \Gamma_{i,j}} \mu^n_{k,i}(x_i) \\ \mu^n_{i,j}(x_j) + \mu^n_{i,j}(\bar{x}_j) - 1. \end{cases} \tag{3}
$$

Let the set of solutions over complex variables, without accounting for multiplicity, be $V(\mathbf{F}) = \{(\underline{\mu},\underline{\alpha}) \in \mathbb{C} : f_m(\underline{\mu},\underline{\alpha}) = 0 \text{ for all } f_m \in \mathbf{F}(\underline{\mu},\underline{\alpha})\}$. In practice we are only interested in strictly positive solutions: $V^*_{\mathbb{R}_+}(\mathbf{F}) \subset V(\mathbf{F})$. If (3) equates to zero, i.e., $(\underline{\mu},\underline{\alpha}) \in V(\mathbf{F})$, if follows that $BP\{\underline{\mu}\} = \underline{\mu}$. Consequently all positive real solutions obtained by solving (3) constitute fixed points of BP.

Whether such systems can be solved in practice depends largely on the chosen method. A great variety of approaches have been developed to solve systems of nonlinear polynomial equation, such as iterative solvers, symbolic methods, and homotopy methods.

One basic method for solving systems of nonlinear equations is Newton's method which is an iterative solver that progressively refines an initial guess to reach a solution. However, this method may diverge or even exhibit chaotic behavior. Moreover, it is difficult, to obtain the full set of solutions. Consequently iterative solvers are not useful in the current setting since we are interested in the entire set of positive real solution $V^*_{\mathbb{R}_+}(\mathbf{F})$. From a completely different point of view, the symbolic methods [4, 5] rely on symbolic manipulation of the equations and successive elimination of variables to obtain a simpler but equivalent form. In a sense, these methods can be considered as various generalizations of the Gaussian elimination method for linear systems into nonlinear settings. In the past several decades, symbolic methods, especially the Gröbner basis method, has seen substantial development. But a worst case complexity that is double exponential in the number of variables [12, 14] and the limited scalability in parallel computation restricts the application to smaller systems.

Another important approach for solving a system of nonlinear equations is the *numerical polynomial homotopy continuation* (NPHC) method. The target system $\mathbf{F}(\underline{\mu},\underline{\alpha})$ in (3), which we intend to solve, is continuously deformed into a closely related start system that is trivial to solve. With an appropriate construction, the corresponding solutions also vary continuously under this deformation forming solution paths that connect the solutions of the start system to the desired solutions of the target system. Each of these solution paths can be tracked independently making this approach being *pleasantly parallelizeable*; this is essential in dealing with large polynomial systems. More details on the NPHC method can be found in [20, 11]

4 Experiments

We apply the NPHC method to obtain all fixed points on fully connected spin glass models. Subsequently we obtain the marginals and compare them with marginals obtained by an implementation of BP without damping [15], as well as with the exact marginals obtained by the junction tree algorithm [10].

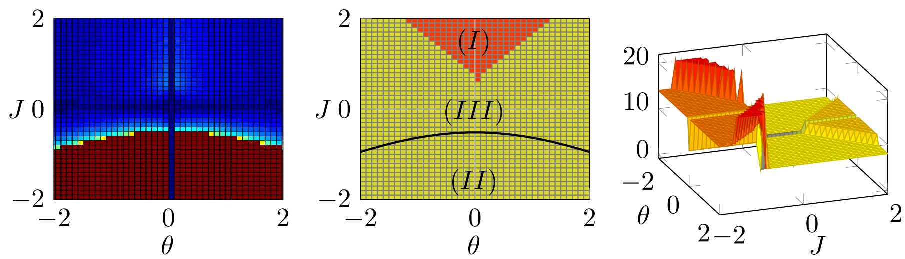
We consider a fully connected Ising model with $|\mathbf{X}| = 4$ RVs. Let the local and pairwise Ising potentials of state x_i be $\Phi_{X_i}(x_i) = \exp(\theta x_i)$ and $\Phi_{X_i,X_j}(x_i,x_j) = \exp(Jx_ix_j)$. Then, the joint distribution is equal to the Boltzmann distribution [13, p.44]:

$$
P(\mathbf{X}=\mathbf{x}) = \frac{1}{Z}\exp\Bigg(\sum_{(i,j)\,:\,e_{i,j}\in\mathbf{E}} Jx_ix_j + \sum_{i=1}^{N}\theta x_i\Bigg). \tag{4}
$$

These models are well studied in the physics literature [6, 13] and exhibit phase transitions, i.e., regions in the parameter-space exist where BP has multiple solutions. In accordance with statistical mechanics we separate the parameter space into three distinct regions (I),(II), and (III) (Fig. 1b). The structure of (3) and the number of complex solutions in $V(F)$ does not change if the underlying structure of the graph is the same [2, 3]; but, depending on the potentials, the number of solutions in $V^*_{\mathbb{R}_+}(\mathbf{F})$ may change. For $(J,\theta) \in (III)$ BP has a unique fixed point, which is a stable attractor

Table 1: MSE OF THE MARGINALS OBTAINED BY BP AND NPHC

Couplings	Local Field	BP	NPHC
$J \in [-2, 2]$	$\theta \in [-2, 2]$	0.069	0.007
$J \in (I)$	$\theta \in (I)$	0.034	0.033
$J \in (II)$	$\theta \in (II)$	0.304	0.004
$J \in (III)$	$\theta \in (III)$	0.003	0.003



(a) (b) (c)

Figure 1: (a) Convergence of BP: for blue BP did converge – for red it did not converge after $4 \cdot 10^5$ iterations. (b) Number of fixed points (yellow: unique fixed point, red: three fixed points). (c) Number of real solutions: at the onset of phase transitions a sudden increase in the number of real solutions can be seen.

in the whole message space [16]. In (I) three fixed points exist, two of which are symmetric, i.e., $\tilde{P}_1(\mathbf{X} = \mathbf{x}_i) = \tilde{P}_2(\mathbf{X} = \bar{\mathbf{x}}_i)$. BP converges to one of these fixed points. If $J < 0$, BP does only converge inside (III) and not inside (II) (Fig. 1a). We can see two interesting effects: first in Fig. 1c the number of real solutions increases at the onset of phase transitions; secondly, even though BP does not converge, a unique fixed point exists inside (II) (Fig. 1b).

We further asses the accuracy of the marginals obtained by NPHC and BP; averaged over all graphs, and for each region (Table 1). Indeed, inside region (II) the marginals obtained by NPHC give much better approximations than BP does.

5 Discussion

The NPHC method is introduced as a tool to obtain all BP fixed point solutions. This work is an attempt to get a deeper understanding of BP, with potential implications for finding stronger conditions for uniqueness of BP fixed points.

First, we create a system of equations which has to be solved in order to obtain all fixed points of BP. Secondly, we present our framework that can be used to obtain the solutions in practice. Finally, we create a simple example and apply the proposed method and show an accuracy-gap between fixed points of BP and the best fixed points obtained by NPHC. In practice this justifies the exploration of multiple fixed points and selecting one that leads to the best approximation.We further show how the number of fixed points evolves over a large parameter region. When applied to graphs where BP does not converge, the NPHC method reveals that for some cases a unique fixed point does exist. While, in practice, fixed points have to be positive, we empirically showed that there is a close relation between the occurrence of phase transitions and an increase in the number of real solutions. In future we aim to exploit the graph structure in order to reduce the complexity of constructing the start system.

Acknowledgments

Research supported in part by NSF under Grant DMS 11-15587. This work was supported by the Austrian Science Fund (FWF) under the project number P7803-N15.

References

[1] D Batra, P Yadollahpour, A Guzman-Rivera, and G Shakhnarovich. Diverse m-best solutions in markov random fields. In *Computer Vision–ECCV 2012*, pages 1–16. Springer, 2012.

[2] T Chen and D Mehta. On the network topology dependent solution count of the algebraic load flow equations. *arXiv preprint arXiv:1512.04987*, 2015.

[3] T Chen, D Mehta, and M Niemerg. A network topology dependent upper bound on the number of equilibria of the kuramoto model. *arXiv preprint arXiv:1603.05905*, 2016.

[4] D Cox, J Little, and D O'shea. *Ideals, varieties, and algorithms*, volume 3. Springer, 1992.

[5] D Cox, J Little, and D O'shea. *Using algebraic geometry*, volume 185. Springer Science & Business Media, 2006.

[6] HO Georgii. *Gibbs Measures and Phase Transitions*, volume 9. 2011.

[7] T Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Comp.*, 16(11), 2004.

[8] Alexander Ihler. Accuracy bounds for belief propagation. In *Proceedings of UAI*, pages 183–190, July 2007.

[9] AT Ihler, J Iii, and AS Willsky. Loopy belief propagation: Convergence and effects of message errors. In *Journal of Machine Learning Research*, pages 905–936, 2005.

[10] SL Lauritzen and DJ Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B*, pages 157–224, 1988.

[11] TY Li. Solving polynomial systems by the homotopy continuation method. *Handbook of numerical analysis*, 11:209–304, 2003.

[12] EW Mayr and AR Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Advances in mathematics*, 46(3):305–329, 1982.

[13] M Mezard and A Montanari. *Information, Physics, and Computation*. Oxford Univ. Press, 2009.

[14] H Möller and F Mora. Upper and lower bounds for the degree of groebner bases. *EUROSAM 84*, pages 172–183, 1984.

[15] JM Mooij. libdai: A free and open source c++ library for discrete approximate inference in graphical models. *The Journal of Machine Learning Research*, 11, 2010.

[16] JM Mooij and HJ Kappen. On the properties of the bethe approximation and loopy belief propagation on binary networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11012, 2005.

[17] JM Mooij and HJ Kappen. Sufficient conditions for convergence of the sum–product algorithm. *IEEE Trans. on Information Theory*, 53(12):4422–4437, 2007.

[18] J Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Representation and Reasoning Series. Morgan Kaufmann Publishers, 1988. ISBN 9781558604797.

[19] F Pernkopf, R Peharz, and S Tschiatschek. *Introduction to Probabilistic Graphical Models*. Academic Press' Library in Signal Processing, 2014.

[20] AJ Sommese and CW Wampler. *The Numerical solution of systems of polynomials arising in engineering and science*, volume 99. World Scientific, 2005.

[21] Christopher Srinivasa, Siamak Ravanbakhsh, and Brendan Frey. Survey propagation beyond constraint satisfaction problems. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 286–295, 2016.

[22] Y Weiss. Correctness of local probability propagation in graphical models with loops. *Neural Comp.*, 12(1), 2000.