Pairwise MRF Models Selection for Traffic Inference

Cyril Furtlehner

Pairwise MRF Models Selection for Traffic Inference

Cyril Furtlehner[*]

Abstract

We survey some recent work where, motivated by traffic inference, we design in parallel two concurrent models, an Ising and a Gaussian ones, with the constraint that they are suitable for "belief-propagation" (BP) based inference. In order to build these model, we study how a Bethe mean-field solution to inverse problems obtained with a maximum spanning tree (MST) of pairwise mutual information, can serve as a reference point for further perturbation procedures. We consider three different ways along this idea: the first one is based on an explicit natural gradient formula; the second one is a link by link construction based on iterative proportional scaling (IPS); the last one relies on a duality transformation leading to a loop correction propagation algorithm on a dual factor graph.

1 Problem statement

Once a joint probability measure is given, the belief propagation algorithm (BP) [12] can be very efficient for inferring hidden variables while observing the others, but in real applications it is often the case that we have first to build the model. This is precisely the case for the application that motivates this work. This deals with the reconstruction and prediction of traffic congestion conditions, typically from sparse observations on the secondary network where no fixed sensors are installed. Data are obtained from vehicles embedded in the traffic, equipped with GPS and able to exchange data through cellular phone connections for example, in the form of so-called Floating Car Data (FCD) by sending their speed along with their position. The goal then is to be able to provide at any time a travel time for each unobserved segment of the network and a short term forecast for all segments.

Our approach to this objective is to build a MRF based on past observations [8]. Each variable representing a travel time is attached to a segment possibly at various discretized time in the day. We assume that the collected data make it possible for a statistical modelling of each segment and a certain number of pairs of segments. The FCD sent by probe vehicles concerning some

[*]Inria Saclay, Bât 660 Université Paris Sud, Orsay Cedex 91405

area of interest are continuously collected over a reasonable period of time (one year or so) such as to allow a finite fraction (a few percents) of road segments to be covered in real time. Schematically the inference method works as follows:

- Historical FCD are used to compute empirical dependencies between contiguous or remote segments of the road network.
- These dependencies are encoded into a graphical model, which vertices are (segment,timestamps) pairs attached with a traffic index variable, like e.g. the binary state CONGESTED/NOT-CONGESTED.
- Congestion probabilities of segments that are unvisited or sit in the short-term future are computed with BP, conditionally to real-time data.

On the factor-graph, the information is propagated both temporally and spatially. In this perspective, reconstruction and prediction are on the same footing, even though prediction is expected to be less precise than reconstruction.

2 MRF models

Since the distribution of travel time tt_ℓ for a given road segment ℓ is not given by a simple parametric and identical model for all segments we do not consider the inference problem in the space of travel time directly but consider instead various mapping $x_\ell(tt_\ell)$ attached to each segment which we call traffic indexes. This leads us to consider basically 2 different models:

- A Gaussian MRF (GMRF): it is the most straightforward approach. First each travel time tt_ℓ is mapped via its empirical cumulative distribution onto a normalized Gaussian variable $x_\ell = \mathcal{N}(0,1)$. The GMRF is then build based on a subset of pairwise empirical covariance $\hat{Cov}(x_\ell, x_{\ell'})$ computed in this space from the observations. In principle BP can be used in closed form in this case, the messages exchanged between variables corresponding to the mean and the variance of these variables.
  Pros: this is well suited real variable inference for BP. If BP converges the marginal are exact marginals of the GMRF model. The inverse mapping directly delivers travel time predictions along with an estimation of the error.
  Cons: there is a strong assumption that the data in the traffic index space have a single mode, which is probably not true for traffic data, where we expect instead to have modes associated to different congestion patterns.
- An Ising model for traffic: this approach is based on the intuition that a natural binary latent state $s_\ell = \pm 1$ for NON-CONGESTED/CONGESTED is attached to each segment. Then, instead of encoding the full dependency between different travel times, we instead use this latent state as a proxy to encode dependencies between segments. Based on the travel time cumulative distribution we propose different ways of defining these latent

state like: (i) a simple fixed threshold with $s_i \stackrel{\text{def}}{=} 2P(tt_\ell < tt^\star_\ell) - 1$ where $tt^\star_\ell$ is e.g. the median (ii) a random threshold allowing one to map the belief $b(s_\ell = 1)$ directly to a travel time tt_ℓ through the inverse cumulative distribution. The correlations between latent states variables needed to feed the MRF model are obtained either by moment matching or via an expectation maximization procedure.
Pros: natural and appealing binary description and a very light interaction model, particularly well suited for BP when a multi-modal distribution is expected to be associated to macro state of congestion patterns.
Cons: variables to predict are real travel time and the mutual information between two travel times is reduced to at most one single bit of mutual information, at least when a fixed threshold (i) is used to define the latent state.

3 The Inverse Problem and an heuristic approximate solution

In both cases we face a difficult inverse problem, where both the MRF's graph structure and parameters have to be determined. In the Ising case these are the local fields and couplings while in the Gaussian case this is the so called "precision matrix" i.e. the inverse covariance matrix which has to be found. In the latter case, there are two issues which prevent the direct use of the inverse covariance matrix: (i) the empirical covariance matrix may not be necessarily full, in our context we expect many missing entries. (ii) the compatibility with BP is not insured i.e. BP might not converge if the precision matrix is too dense. Therefore in the Gaussian case we look for a good trade-off between likelihood and sparsity of the model.

A statistical physics approach to the inverse Ising problem (IIP) is given by the linear response theory combined with various hypothesis. It is usually based on the Plefka expansion [13] of the Gibbs free energy by making the assumption that the coupling constants J_{ij} are small. At lowest order the perturbation expansion delivers the mean field solution, the TAP solution at the next order …A different type of mean-field approximation is the Bethe approximation which reduces to the TAP approximation at lowest order and which consists in to assume that relevant coupling J_{ij} have locally a tree-like structure. The Bethe approximation [16] can then be used and leads to approaches referred to as pseudo-moment matching methods [9, 14, 10, 15]. This basically leads to two, possibly different mean-field solutions to the IIP: the direct one, by using the relation valid on a tree between the joint probability and the single and pairwise marginal distributions; the indirect one also called susceptibility propagation [10] relying in fact on the relation between the inverse susceptibility matrix and the set of susceptibility coefficients attached to the links of the tree [11].

These approaches are mainly valid if we expect a high temperature model in its paramagnetic phase. Instead our data displays low temperature behaviour and cannot be considered as uni-modal. A classical results of [3] concerning inference using dependence trees, states that given a set of single and pairwise marginals $p_i(x_i)$ and $p_{ij}(x_i, x_j)$, an optimal model can be obtained on the subclass of singly connected graphical model by considering a MST $\mathcal{T}$ w.r.t. mutual information, with joint measure

$$\mathcal{P}(\mathbf{x}) = \prod_{(ij)\in\mathcal{T}} \frac{p_{ij}(x_i, x_j)}{p_i(x_i)p_j(x_j)} \prod_{i\in\mathcal{V}} p_i(x_i). \tag{1}$$

This suggests it is a good starting point from which we should perturb the Bethe approximation to find improved solutions compatible with BP. We refer to as the *Bethe reference point*, the approximate solution (1). In [7] we have proposed a simple heuristic 2-parameters deformation of this model, with K the mean connectivity of the graph and α a global rescaling of the interactions. The graph is obtained by completing the MST up to the targeted mean connectivity K by selecting links according to the mutual information they carry. The parameter $\alpha \in [0, 1]$ allows one to interpolate between a model with independent variables with single marginal matching the observation for $\alpha = 0$, and the direct Bethe solution for $\alpha = 1$. Upon joint calibration of α and K, the model is able to recover well separated modes of a multimodal distribution. In that case, the model displays many BP fixed points, in one to one correspondence with the hidden modes contained in the observation data. We can interpret this heuristic solution through some asymptotic mapping onto a Hopfield model.

4 Various Perturbations of the Bethe reference point

To go beyond this simple heuristic, we have looked for some theoretic foundation for deforming this Bethe reference point. We have actually studied three different but complementary ways of proceeding [6].

4.1 Line search along the natural gradient direction

A first direction is provided by the observation that the natural gradient can be made explicit at the Bethe point. We develop it for the inverse Ising problem. What we propose amounts simply to change the reference point for the Plefka's expansion of the Gibbs free energy, i.e. to use the Bethe reference point instead of the free uncorrelated one. The set of couplings are then decomposed into $J_{ij}^{Bethe} + J_{ij}$ where J_{ij}^{Bethe} are the Bethe reference point couplings attached to the MST and J_{ij} are considered to be small. The observations consist of empirical single and pairwise expectations $\hat{m}_i = \hat{\mathbb{E}}(s_i)$ and $\hat{m}_{ij} = \hat{\mathbb{E}}(s_i s_j)$ and the Bethe susceptibility matrix $[\chi_{Bethe}]$ is assumed to be given. The Gibbs free energy $G[\mathbf{J}]$ then reads at second order

$$G[J] = G_{Bethe} + G_{BLR}[\mathbf{J}] + o(\mathbf{J}^2).$$

with

$$G_{BLR}[\mathbf{J}] \stackrel{\text{def}}{=} \mathbb{E}_{Bethe}(H^1) + \frac{1}{2}\Big(\mathrm{Var}_{Bethe}(H^1) - \sum_{i,j} [\chi_{Bethe}^{-1}]_{ij} \,\mathrm{Cov}_{Bethe}(H^1, s_i)\,\mathrm{Cov}_{Bethe}(H^1, s_j)\Big),$$

$H^1 = \sum_{ij} J_{ij} s_i s_j$ being the perturbation, while all quantities with subscript $_{Bethe}$ are evaluated using the reference measure (1). They involve up to 4-point susceptibility coefficients, explicitly given in [6] in terms of the Bethe susceptibility coefficients. Concerning the log-likelihood, it is given now by:

$$\mathcal{L}[\mathbf{J}] = -G_{Bethe} - G_{BLR}[\mathbf{J}] - \sum_{ij} (J_{ij}^{Bethe} + J_{ij})\hat{m}_{ij} + o(J^2).$$

G_{BLR} is at most quadratic in the J's and contains the local projected Hessian of the log likelihood onto the magnetization constraints with respect to this set of parameters. This is nothing else than the Fisher information matrix associated to these parameter J. Therefore it makes sense to use this quadratic approximation to find the optimal point. Given a parametrization of the deformation H^1, the natural gradient can then be made explicit and tractable optimization strategies like e.g. a line-search along this direction may be implemented.

4.2 Iterative proportional scaling

A second direction that we have explored consists in to proceed link-wise from the Bethe reference point. The link yielding the maximum gain in likelihood is obtained by solving a simple variational problem which solution is referred to as "iterative proportional scaling" (IPS) in the statistics literature, for solving maximum likelihood estimation problem [4, 5].

Suppose we start from a graphical model based on some graph $\mathcal{G}^{(0)}$ and we want to add one link to it to obtain a graph $\mathcal{G}^{(1)}$. Letting $\mathcal{P}^{(0)}$ be the reference distribution to which we want to add one factor ψ_{ij} to produce the distribution

$$\mathcal{P}^{(1)}(\mathbf{x}) = \mathcal{P}^{(0)}(\mathbf{x}) \times \frac{\psi_{ij}(x_i,x_j)}{Z_\psi} \qquad \text{with} \qquad Z_\psi = \int dx_i dx_j p_{ij}^{(0)}(x_i,x_j)\psi_{ij}(x_i,x_j). \tag{2}$$

The max log likelihood corresponding to this new distribution is obtained for

$$\psi_{ij}(x_i,x_j) = \frac{\hat{p}_{ij}(x_i,x_j)}{p_{ij}^{(0)}(x_i,x_j)} \qquad \text{with} \qquad Z_\psi = 1,$$

where $p^{(0)}(x_i,x_j)$ is the reference marginal distribution obtained from $\mathcal{P}^{(0)}$. The log-likelihood increase reads then

$$\Delta\mathcal{L} = D_{KL}(\hat{p}_{ij} \| p_{ij}^{(0)}). \tag{3}$$

Sorting all the links w.r.t. this quantity yields the (exact) optimal 1-link correction to be made. The interpretation is therefore immediate: the best candidate is the one for which the current model yields the joint marginal $p_{ij}^{(0)}$ that is most distant from its target, i.e. the empirical marginal $\hat{p}_{ij}$. Note that the update mechanism is indifferent to whether the link has to be added or simply modified.

The algorithm that we propose in [6] is then to start from the Bethe reference point based on the maximum spanning tree, and to complete the graph link by link by this selection mechanism up to some mean connectivity. At each step it is needed to compute the covariance matrix, in order to be able to rank the candidate link with the criteria given in (3).

In the GMRF case we found that this can be implemented efficiently due to local transformations of the precision matrix after adding one link. By comparison we implemented methods based on sparse norm penalization (L_0 or L_1) and find out that it is competitive with L_0 based approach, with a $O(N^3)$ complexity in the sparse regime. Incidentally we also found that the L_1 based method is not working well for this problem. An additional advantage of our IPS based solution is the possibility to combine it with spectral constraints like walk-summability with BP or/and graph structure constraint to enforce compatibility. For the same computational cost we can get a complete set of good trade-off between likelihood and compatibility with BP.

Concerning the Ising model IPS is too computationally expensive, and even if we separate the structure from the coupling selection, there is no satisfactory solution in the low temperature regime. It can be used only for marginal modification of a Bethe model.

4.3 Duality transformation and loop corrections

For the Ising model a standard way to deal with the high couplings at low temperature is provided by a duality transformation, which in absence of local fields leads to a dual model of binary loop variables with weak interactions. When looking at the explicit formula for the Bethe susceptibility we can see that it potentially incorporates loop corrections which are wrong already at the first loop contribution. This explains why susceptibility propagation is not working well in this domain. Our proposal in this context to provide approximate solutions to the IIP in a tractable way, is to use a loop variables joint model. This is obtained in absence of external fields by rewriting the factors of the Gibbs measure as

$$e^{J_{ij}s_is_j} = \cosh(J_{ij})(1 + \tanh(J_{ij})s_is_j). \tag{4}$$

Using this identity the partition function rewrites

$$Z(\mathbf{J}) = Z_0 \times \sum_{\{\tau_{ij}\in\{0,1\}\}} \prod_{(ij)\in\mathcal{E}} \bigl(\bar{\tau}_{ij} + \tau_{ij}\tanh(J_{ij})\bigr) \times \prod_i \mathbb{1}_{\{\sum_{j\in\partial i}\tau_{ij}=0 \mod 2\}},$$

with

$$Z_0 = \prod_{(ij)} \cosh(J_{ij}).$$

The summation over bond variables $\tau_{ij} \in \{0,1\}$ ($\bar{\tau}_{ij} \stackrel{\text{def}}{=} 1 - \tau_{ij}$), corresponds to choosing one of the 2 terms in the factor (4). The summation over spin variables then selects bond configurations having an even number of bonds $\tau_{ij} = 1$ attached to each vertex i. From this condition it results that the paths formed by these bonds must be closed. The contribution of a given path is simply the product of all bond factor $\tanh(J_{ij})$ along the path. As a result, the partition function is expressed as

$$Z(\mathbf{J}) = Z_0 \times Z_{loops} \qquad \text{with} \qquad Z_{loops} \stackrel{\text{def}}{=} \sum_{\ell} Q_{\ell},$$

where the last sum runs over all possible closed loops $\mathcal{G}_\ell$, i.e. subgraphs for which each vertex has an even degree, including the empty graph and

$$Q_\ell \stackrel{\text{def}}{=} \prod_{(ij)\in\mathcal{E}_\ell} \tanh(J_{ij}),$$

where $\mathcal{E}_\ell$ denotes the set of edges involved in loop $\mathcal{G}_\ell$. This is a special case of the loop expansion around a belief propagation fixed point proposed by Chertkov and Chernyak in [2]. When there are local fields, variables can be approximately centered with help of a BP fixed point, Z_0 is to be replaced by the associated Z_{BP} and the loop corrections runs over all generalized loops, i.e. all subgraphs containing no vertex with degree 1.

In absence of external field, loops which contribute have a simpler combinatorial structure. If the graph has $k(\mathcal{G})$ connected components, we may define a set $\{\mathcal{G}_c, c = 1, \dots, C(\mathcal{G})\}$ of independent cycles, $C(\mathcal{G}) = |\mathcal{E}| - |\mathcal{V}| + k(\mathcal{G})$ being the so-called cyclomatic number [1] of graph $\mathcal{G}$. Spanning the set $\{0,1\}^{C(\mathcal{G})}$ yields all possible loops with the convention that edges are counted modulo 2 for a given cycle superposition. The partition function can therefore be written as a sum over dual binary variables $\tau_c \in \{0,1\}$ attached to each cycle $c \in \{1, \dots |C|\}$:

$$Z_{loops} = \sum_{\tau} Q_{\mathcal{G}^\star}(\tau), \tag{5}$$

where $Q_{\mathcal{G}^\star}(\tau)$ represents the weight for any loop configuration specified by $\{\tau_c\}$ on the dual (factor-)graph $\mathcal{G}^\star$ formed by the cycles. For instance, when the primal graph $\mathcal{G}$ is a 2-d lattice, the dual one is also 2-d and the Kramers-Wannier duality expresses the partition function at the dual coupling $J^\star \stackrel{\text{def}}{=} -\frac{1}{2}\log(\tanh(J))$ of the associated Ising model on this graph, with spin variable $\sigma_c = 2\tau_c - 1$ attached to each plaquette representing an independent cycle c. Let

$$Q_c \stackrel{\text{def}}{=} \prod_{(ij)\in\mathcal{E}_c} \tanh(J_{ij}),$$

be the free weight attached to each cycle c. Assuming that there exists a basis such that each link belongs to at most 2 cycles and each cycle has a link in common with at most one other cycle, the partition function then reads

$$Z_{loops} = \sum_{\tau} \prod_{c=1}^{C(\mathcal{G})} (\bar{\tau}_c + \tau_c Q_c) \prod_{c,c'} \big(\bar{\tau}_c\bar{\tau}_{c'} + \tau_c\bar{\tau}_{c'} + \bar{\tau}_c\tau_{c'} + \tau_c\tau_{c'}Q_{cc'}\big)$$

where

$$Q_{cc'} \stackrel{\text{def}}{=} \Big(\prod_{(ij)\in\mathcal{E}_c\cap\mathcal{E}_{c'}} \tanh(J_{ij})\Big)^{-2}.$$

Since the sign of $Q_{\mathcal{G}^\star}(\tau)$ is not guaranteed to be positive, there is possibly no probability interpretation for these weights. Nevertheless, if the dual graph $\mathcal{G}^\star$ is singly connected, we can proceed analogously to ordinary belief propagation and set up an exact message passing procedure to compute the following weights:

$$q_c \stackrel{\text{def}}{=} \frac{1}{Z_{loops}} \sum_\tau \tau_c Q_{\mathcal{G}^\star}(\tau), \qquad q_{cc'} \stackrel{\text{def}}{=} \frac{1}{Z_{loops}} \sum_\tau \tau_c \tau_{c'} Q_{\mathcal{G}^\star}(\tau).$$

The susceptibility coefficients can then be expressed in terms of these weights

$$\chi_{ij} = \tanh(J_{ij}) + \frac{1-\tanh^2(J_{ij})}{\tanh(J_{ij})}\Big(\sum_{\substack{c,\\(ij)\in\mathcal{G}_c}} q_c - 2\sum_{\substack{cc',\\(ij)\in\mathcal{E}_c\cap\mathcal{E}_{c'}}} q_{cc'}\Big).$$

The weights q_c and $q_{cc'}$ can be computed as follows, by "loop weight propagation". The message passing procedure involves messages of the form

$$m_{c'\to c}(\tau_c) = (1-m_{c'\to c})\bar{\tau}_c + m_{c'\to c}\tau_c,$$

which update rules are given by

$$m_{c\to c'} = \frac{1+r_{c\to c'}Q_cQ_{cc'}}{2+r_{c\to c'}Q_c(1+Q_{cc'})}, \qquad \text{where} \qquad r_{c\to c'} = \prod_{c''\in\partial c\setminus c'} \frac{m_{c''\to c}}{1-m_{c''\to c}},$$

∂c representing the neighborhood of c in $\mathcal{G}^\star$. Finally, letting

$$\nu_{c\to c'} \stackrel{\text{def}}{=} \frac{m_{c\to c'}}{1-m_{c\to c'}},$$

leads to the following loop weights propagation update rules:

$$\begin{cases} \nu_{c\to c'} \longleftarrow \dfrac{1+r_{c\to c'}Q_cQ_{cc'}}{1+r_{c'\to c}Q_c}, \\ r_{c\to c'} \longleftarrow \displaystyle\prod_{c''\in\partial c\setminus c'} \nu_{c''\to c}. \end{cases}$$

When a fixed point is reached, we obtain from these messages the following expressions for the cycle weights:

$$q_c = \frac{Q_cr_c}{1+Q_cr_c} \qquad \text{and} \qquad q_{cc'} = \frac{Q_cQ_{c'}Q_{cc'}r_{c'\to c}r_{c\to c'}}{1+Q_cr_{c\to c'}+Q_{c'}r_{c'\to c}+Q_cQ_{c'}Q_{cc'}r_{c'\to c}r_{c\to c'}},$$

with

$$r_c \stackrel{\text{def}}{=} \prod_{c'\in\partial c} \nu_{c'\to c}.$$

Another useful expression resulting from the message passing machinery, is the possibility to express the partition function in terms of the single and pairwise weights normalizations. Introducing also

$$s_c \stackrel{\text{def}}{=} \prod_{c'\in\partial c}(1+\nu_{c'\to c}), \qquad s_{c'\to c} \stackrel{\text{def}}{=} \prod_{c''\in\partial c'\setminus c}(1+\nu_{c''\to c'}),$$

we have

$$Z_{loops} = \prod_{c\in\mathcal{V}^\star} Z_c \prod_{(cc')\in\mathcal{E}^\star} \frac{Z_{cc'}}{Z_c Z_{c'}}.$$

with

$$Z_c = \frac{1+Q_c r_c}{s_c} \qquad \text{and} \qquad Z_{cc'} = \frac{1+Q_c r_{c\to c'} + Q_{c'} r_{c'\to c} + Q_c Q_{c'} Q_{cc'} r_{c'\to c} r_{c\to c'}}{s_{c\to c'} s_{c'\to c}}.$$

In [6] we actually show that this dual formulation and associated message passing algorithm can be extended to any dual-loop free graph, not necessarily restricted to pairwise dual factor graph and constitute sometimes a good approximation when $\mathcal{G}^\star$ has loops. In turn this formulation can be used to determine the Ising couplings when the graph structure is given and could be well combined with IPS when this is not the case, the linear response being now given explicitly.

5 Conclusion

We underline in this short survey that using the Bethe approximation as a starting point for further corrections can be fruitful, in particular for building inverse models from data observations. We have exposed here three different ways of perturbing such a mean-field solution. The IPS based method is valid both for binary and Gaussian variables and leads to an efficient algorithm in the Gaussian case to generated sparse approximation models compatible with BP. Some ongoing work is devoted to extend the two other methods. In particular the loop correction propagation should not be limited to the Ising model with no external field, and using the "cycle-tree" property by analogy with the definition of junction-tree as a graph hypothesis leads actually to formulate loop correction propagation algorithms for general models (work in progress).

Acknowledgements: The work summarized here was done in collaboration with Jean-Marc Lasgouttes, Yufei Han and Victorin Martin.

References

[1] Berge, C. *Théorie des graphes et ses applications*, 2ème ed., vol. II of *Collection Universitaire des Mathématiques*. Dunod, 1967.

[2] Chertkov, M., and Chernyak, V. Y. Loop series for discrete statistical models on graphs. *J.Stat.Mech.* (2006), P06009.

[3] Chow, C., and Liu, C. Approximating discrete probability distributions with dependence trees. *Information Theory, IEEE Transactions on 14*, 3 (1968), 462 – 467.

[4] Darroch, J., and Ratcliff, D. Generalized iterative scaling for log-linear models. *Ann. Math. Statistics 43* (1972), 1470–1480.

[5] Della Pietra, S., Della Pietra, V., and Lafferty, J. Inducing features of random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 19*, 4 (1997), 380 –393.

[6] Furtlehner, C., Han, Y., Lasgouttes, J.-M., and Martin, V. Pairwise MRF Calibration by Perturbation of the Bethe Reference Point. *ArXiv e-prints* (2012).

[7] Furtlehner, C., Lasgouttes, J.-M., and Auger, A. Learning multiple belief propagation fixed points for real time inference. *Physica A: Statistical Mechanics and its Applications 389*, 1 (2010), 149–163.

[8] Furtlehner, C., Lasgouttes, J.-M., and de La Fortelle, A. A belief propagation approach to traffic prediction using probe vehicles. In *Proc. IEEE 10th Int. Conf. Intel. Trans. Sys.* (2007), pp. 1022–1027.

[9] Kappen, H., and Rodríguez, F. Efficient learning in Boltzmann machines using linear response theory. *Neural Computation 10*, 5 (1998), 1137–1156.

[10] Mézard, M., and Mora, T. Constraint satisfaction problems and neural networks: A statistical physics perspective. *Journal of Physiology-Paris 103*, 1-2 (2009), 107 – 113.

[11] Nguyen, H., and Berg, J. Bethe-Peierls approximation and the inverse Ising model. *J. Stat. Mech.*, 1112.3501 (2012), P03004.

[12] Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*. Morgan Kaufmann, 1988.

[13] Plefka, T. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *J. Phys. A: Mathematical and General 15*, 6 (1982), 1971.

[14] Welling, M., and Teh, Y. Approximate inference in Boltzmann machines. *Artif. Intell. 143*, 1 (2003), 19–50.

[15] Yasuda, M., and Tanaka, K. Approximate learning algorithm in Boltzmann machines. *Neural Comput. 21* (2009), 3130–3178.

[16] Yedidia, J. S., Freeman, W. T., and Weiss, Y. Generalized belief propagation. *Advances in Neural Information Processing Systems* (2001), 689–695.