Fast Convergence of Belief Propagation to Global Optima: Beyond Correlation Decay

Frederic Koehler

Massachusetts Institute of Technology

Slides for NeurIPS 2019
Inference in graphical models

Ising model:

$$\Pr(X = x) = \frac{1}{Z} \exp \left( \frac{1}{2} x^T J x + h^T x \right)$$

Natural model of correlated random variables. Some examples: Hopfield networks, Restricted Boltzmann Machine (RBM) = bipartite Ising model
**Inference**: Given $J, h$ compute properties of the model. E.g. $E[X_i]$ or $E[X_i | X_j = x_j]$, etc. Can largely be reduced to estimating $\log Z$ (look at derivatives).

![Graph with nodes $X_a$, $X_b$, $X_c$, $X_e$, $X_f$ connected]

$
E[X_e | X_a = 1] = ?
$

**Problem**: inference in Ising models (e.g. approximating $E[X_i]$) is NP-hard! Natural markov chain approaches (e.g. Gibbs sampling) may mix very slowly.
Message passing algorithms

A major approach to inference: variational methods + message-passing algorithms. Deterministic and often faster than MCMC.

Mean-field iteration:

\[ x^{(t+1)} = \tanh^{\otimes n}(Jx^{(t)} + h) \]

Belief propagation:

\[ \nu^{(t+1)}_{i \rightarrow j} = \tanh \left( h_i + \sum_{k \in \partial i \setminus j} \tanh^{-1}(\tanh(J_{ik})\nu^{(t)}_{k \rightarrow i}) \right) \]

These are specialized optimization algorithms attempting to solve an variational problem which approximates the true Ising model by a simpler (pseudo-)distribution.
Our Assumption

We suppose that

\[ J_{ij} \geq 0, h_i \geq 0 \]

for all \( i, j \). This is referred to as **ferromagneticity**, it means the Ising model is attractive in the sense that neighboring spins wants to align. (Natural for modeling social networks, etc.)

This assumption is **necessary**: if we don’t have it, computing the optimal mean-field approximation, even approximately, is NP hard. (By reduction to MAX-CUT).

Under only this assumption, we show that from all-1s initialization the message passing algorithms do indeed converge (quickly) to global optima. **Initialization matters!** Convergence slow/fails from other points.
Our Theorems

Fix a ferromagnetic Ising model \((J, h)\) with \(m\) edges and \(n\) nodes.

**Theorem (Mean-Field Convergence)**

Let \(x^*\) be a global maximizer of \(\Phi_{MF}\). Initializing with \(x^{(0)} = \vec{1}\) and defining \(x^{(1)}, x^{(2)}, \ldots\) by iterating the mean-field equations, for every \(t \geq 1\):

\[
0 \leq \Phi_{MF}(x^*) - \Phi_{MF}(x^{(t)}) \leq \min \left\{ \frac{\|J\|_1 + \|h\|_1}{t}, 2 \left( \frac{\|J\|_1 + \|h\|_1}{t} \right)^{4/3} \right\}.
\]

**Theorem (BP Convergence)**

Let \(P^*\) be a global maximizer of \(\Phi_{Bethe}\). Initializing \(\nu_{i \rightarrow j}^{(0)} = 1\) for all \(i \sim j\) and defining \(\nu^{(1)}, \nu^{(2)}, \ldots\) by BP iteration,

\[
0 \leq \Phi_{Bethe}(P^*) - \Phi_{Bethe}^*(\nu^{(t)}) \leq \sqrt{\frac{8mn(1 + \|J\|_\infty)}{t}}.
\]
For More

The poster: Poster 174, Wednesday 10:45-12:45
The paper: https://arxiv.org/abs/1905.09992