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

Frederic Koehler

Massachusetts Institute of Technology

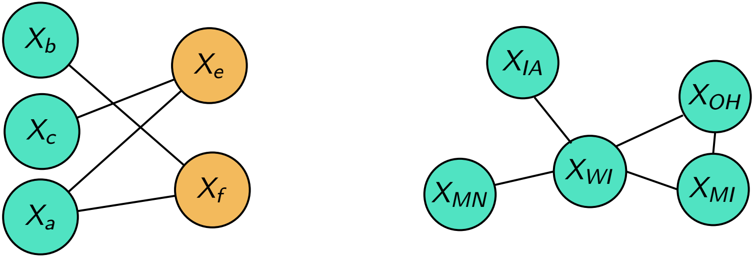
NeurIPS 2019

Graphical models

Ising model: For $x \in \{\pm 1\}^n$,

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

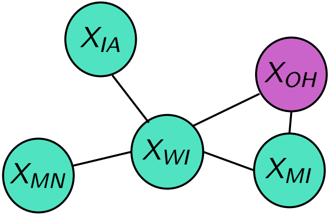
Natural model of correlated random variables. Some examples: Hopfield networks, Restricted Boltzmann Machine (RBM) = bipartite Ising model.



Popular model in ML, natural and social sciences, etc.

Inference

Inference: Given J, h compute properties of the model. E.g. $\mathbb{E}[X_i]$ or $\mathbb{E}[X_i | X_j = x_j]$.



$$\mathbb{E}[X_{WI} \mid X_{OH} = +1] = ?$$

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

Variational Inference

Variational objectives (Mean-Field/VI, Bethe):

$$\Phi_{MF}(x) = \frac{1}{2}x^T Jx + h^T x + \sum_i H\left(Ber\left(\frac{1+x_i}{2}\right)\right)$$

$$\Phi_{Bethe}(P) = \mathbb{E}_P[\frac{1}{2}X^T JX + h^T X] + \sum_E H_P(X_E) - \sum_i (\deg(i) - 1)H_P(X_i)$$

Message-passing algorithms (MF/VI, BP):

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

$$\nu_{i\to j}^{(t+1)} = \tanh\left(h_i + \sum_{k\in\partial i\setminus j} \tanh^{-1}(\tanh(J_{ik})\nu_{k\to i}^{(t)})\right)$$

Non-convex objective — when do these algorithms find global optima?

Our Assumption

We suppose, following Dembo-Montanari '10, that the model is **ferromagnetic**:

$$J_{ij} \geq 0, h_i \geq 0 \quad \text{for all } i, j$$

- I.e. neighbors want to align.
- This assumption is **necessary**: if we don't have it, computing the optimal mean-field approximation, even approximately, is NP hard.
- Objective typically has sub-optimal critical points. (cf. correlation decay)

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 Φ_{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 Φ_{Bethe}. Initializing $\nu^{(0)}_{i\to j} = 1$ 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`