Local Message Passing on Frustrated Systems

Luca Schmid, Joshua Brenk and Laurent Schmalen

Marginal Inference

Probabilistic Inference

Infer the state of a system $\mathcal{X} = \{x_1, \dots, x_N\}$ based on a noisy observation $\boldsymbol{y}$ and prior knowledge $p(\mathcal{X})$.

- Goal: Find the corresponding **posterior distribution** $p(\mathcal{X}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\mathcal{X}) \cdot p(\mathcal{X})}{p(\boldsymbol{y})}$
- **Marginal inference**: Infer the state of a **local** variable x_n
  - **Marginalization**: $P(x_n|\boldsymbol{y}) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} P(\mathcal{X}|\boldsymbol{y}) =: \sum_{\sim\{x_n\}} P(\mathcal{X}|\boldsymbol{y})$

→ computationally complex for large systems $\mathcal{X}$

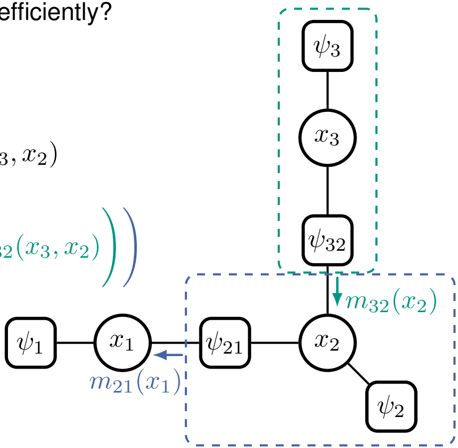
Marginal Inference on Graphical Models

- Assume structure $p(\mathcal{X}|\boldsymbol{y}) = \frac{1}{Z}\prod_{n=1}^{N}\psi_n(x_n)\prod_{n>m}\psi_{nm}(x_m, x_n)$
- Can we use this structure to perform the marginalization more efficiently?
  → **distributive law** $\widehat{=}$ **message passing on graphs**

Example:

$$p(x_1|\boldsymbol{y}) = \frac{1}{Z}\sum_{x_2}\sum_{x_3}\psi_1(x_1)\psi_2(x_2)\psi_3(x_3)\psi_{21}(x_2,x_1)\psi_{32}(x_3,x_2)$$

$$= \frac{1}{Z}\psi_1(x_1)\left(\sum_{x_2}\psi_2(x_2)\psi_{21}(x_2,x_1)\left(\sum_{x_3}\psi_3(x_3)\psi_{32}(x_3,x_2)\right)\right)$$

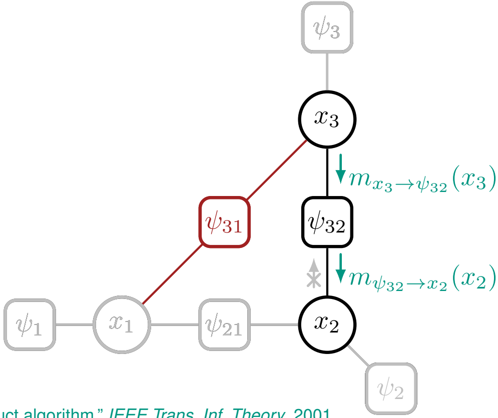


Message Passing on Graphical Models

Sum-product Algorithm (SPA) [KFL01]

$$m_{\psi \to x}(x) = \sum_{\sim\{x\}} \left(\psi(\mathcal{X}_n) \prod_{x' \in \mathcal{X}_n \setminus \{x\}} m_{x' \to \psi}(x') \right)$$

$\mathcal{X}_n \subset \mathcal{X}$: Adjacent variable nodes of factor node $\psi(\mathcal{X}_n)$



Advantages of message passing on graphs:

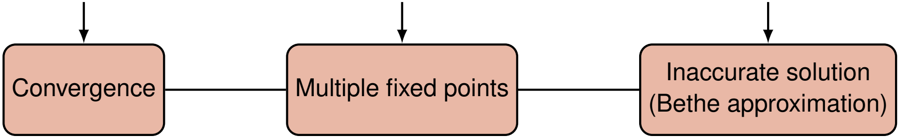
- Algorithm defined by **simple update rule**
- **Efficient algorithms** can be derived very intuitively
- Also **applicable to graphs with cycles**

[KFL01] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, 2001.

Message Passing on Graphs with Cycles

Karlsruher Institut für Technologie

- SPA also applicable to cyclic graphs → iterative algorithm
- **But:** SPA only yields **approximation of marginals** on graphs with cycles
  - **Fixed points** of SPA correspond to **stationary points** of the **Bethe free energy** **[YFW00]**



- Alternative: Directly try to minimize the Bethe free energy, e.g., concave-convex procedure (CCCP) **[Yui02]**
- Many attemps to **mitigate this suboptimality of the SPA** on cyclic graphs: Momentum BP, Neural BP, . . .

→ Instead of fixing a mismatched algorithm, we want to
find a message update rule that is especially taylored to graphs with (many) cycles!

[YFW00] J. S. Yedidia, W. T. Freeman, and Y. Weiss. "Generalized belief propagation," *Advances in Neural Information Processing Systems*, 2000.
[Yui02] A. L. Yuille. "CCCP algorithms to minimize the Bethe and Kikuchi free energies," *Neural Computation*, 2002.

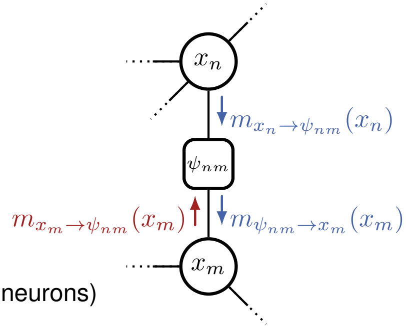
Message Passing on Graphs with Cycles

Message Update Rule

A mapping from one or multiple incident messages to one outgoing message, which is applied **locally** at a (factor) node.

Extrinsic: $\text{FN}_\text{e}(\psi_{n,m}) : m_{x_n \to \psi_{n,m}}(x_n) \mapsto m_{\psi_{n,m} \to x_m}(x_m)$

Non-extrinsic: $\text{FN}(\psi_{n,m}) : \begin{pmatrix} m_{x_n \to \psi_{n,m}}(x_n) \\ m_{x_m \to \psi_{n,m}}(x_m) \end{pmatrix} \mapsto m_{\psi_{n,m} \to x_m}(x_m)$



- Represent mapping by compact multilayer perceptron (1 hidden layer à 7 neurons)
  - **cycBP: Optimize local mapping towards good end-to-end inference performance**

Marginal Inference: Objective Functions

- Supervised Training: **Kullback-Leibler (KL) divergence** $\mathcal{L}_{\text{KL}} := D_{\text{KL}}\left(b_n(x_n) \| p(x_n)\right)$
  - $b_n(x_n), b_{n,m}(x_n, x_m)$: Single, pairwise beliefs of message passing
  - $p(x_n) = \sum_{\sim\{x_n\}} p(x_1, \dots, x_N)$: Exact marginal distributions
- Unsupervised Training: **Regularized Bethe free energy** $\mathcal{L}_{\text{Bethe}} := F_{\text{Bethe}} + \alpha \mathcal{L}_{\mathbb{L}}$ $, \alpha \in \mathbb{R}^+$
  - Bethe free energy $F_{\text{Bethe}} = \sum_{(n,m)\in\mathcal{E}} \sum_{x_n,x_m} b_{n,m}(x_n, x_m) \log\left(\frac{b_{n,m}(x_n,x_m)}{\psi_n(x_n)\psi_{n,m}(x_n,x_m)\psi_m(x_m)}\right)$

$$- \sum_{n=1}^{N} (|\mathcal{X}_n| - 1) \sum_{x_n} b_n(x_n) \log\left(\frac{b_n(x_n)}{\psi_n(x_n)}\right)$$

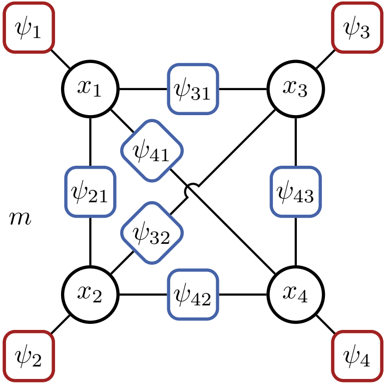
  - **Bethe consistency distance** (*proposed*)

$$\mathcal{L}_{\mathbb{L}} := D_{\text{KL}}\left(\sum_{x_m} b_{n,m}(x_n, x_m) \middle\| b_n(x_n)\right) + D_{\text{KL}}\left(\sum_{x_n} b_{n,m}(x_n, x_m) \middle\| b_m(x_m)\right)$$

Experiment: 2×2 Ising Model

Karlsruher Institut für Technologie

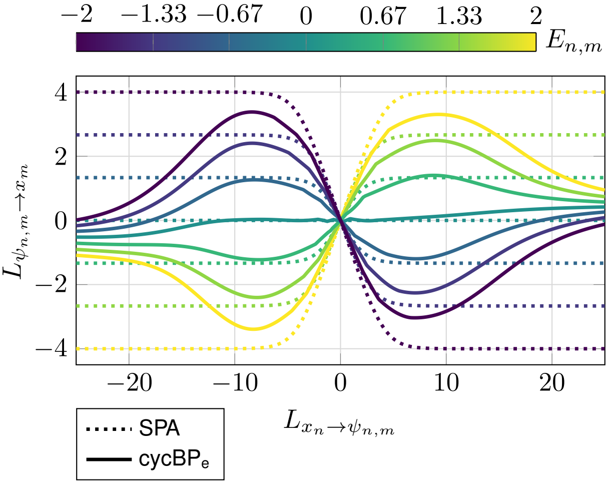
- $N = 4$ binary variables $x_n \in \{+1, -1\}$
  - Use log-likelihood ratios $L_{\psi_{nm}\to x_n} := \log\left(\frac{m_{\psi_{nm}\to x_n}(x_n=+1)}{m_{\psi_{nm}\to x_n}(x_n=-1)}\right)$
- $\psi_n(x_n) = \exp(\theta_n x_n)$ with **local fields θ_n**, $n = 1, \dots, N$
- $\psi_{n,m}(x_n, x_m) = \exp(E_{n,m} x_n x_m)$ with **local couplings $E_{n,m}$**, $n > m$
- **Spin glass**: θ_n and $E_{n,m}$ independently sampled from $\mathcal{U}[-2, +2]$
  - **Frustrated system**, e.g., for $E_{n,m} \ll 0, \forall n, m$



Experiment: 2×2 Ising Model

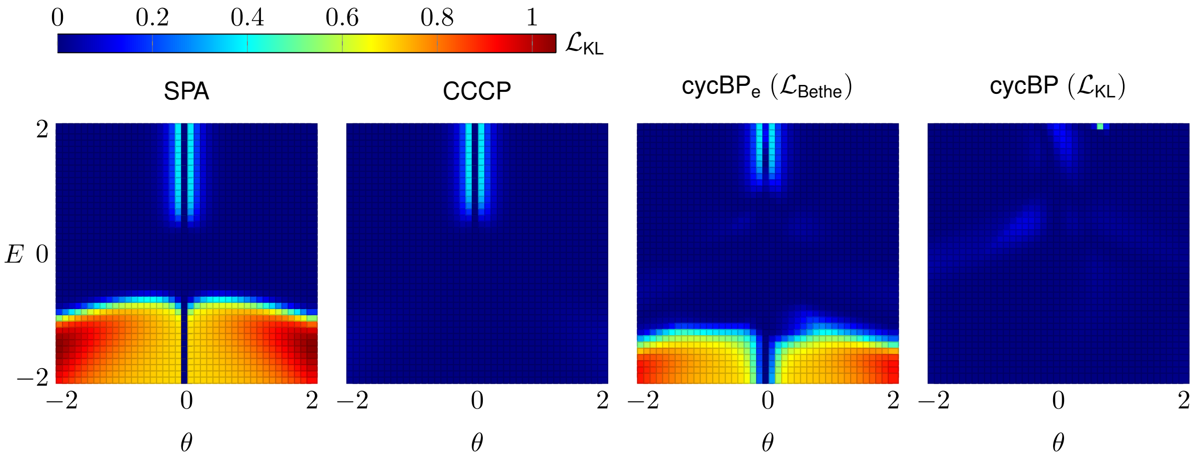
- **Training** and **evaluation** (averaged over 10^5 graphs) on samples of the Ising spin glass model

Algo.	Loss	$\mathcal{L}_{\text{KL}}$	$\mathcal{L}_{\mathbb{L}}$
SPA	-	0.087	0.30
CCCP	-	0.044	$2 \cdot 10^{-6}$
cycBP_e	$\mathcal{L}_{\text{KL}}$	0.040	0.17
cycBP_e	$\mathcal{L}_{\text{Bethe}}$	0.030	0.11
cycBP	$\mathcal{L}_{\text{KL}}$	0.014	0.48
cycBP	$\mathcal{L}_{\text{Bethe}}$	0.027	0.027



Experiment: 2×2 Ising Model

- **Training** on samples of the Ising spin glass model (as before)
- **Evaluate generalization capability** on graphs with constant parameters $E_{nm} = E$ and $\theta_n = \theta$



Conclusion & Outlook

- Alternative message update rules exist which perform especially well on cyclic graphs where the SPA fails
- **Extrinsic principle loses its objective** on graphs with many short cycles
- **Regularized Bethe free energy $\mathcal{L}_{\text{Bethe}}$** as novel objective function for unsupervised training
- Results for **symbol detection** as a practical application → **Poster session!**

- Investigate optimal message update rules for larger graphs and different topologies
- Replace NN by simple parametric function
- Make update rules a function of the node degree, the iteration number, . . .

This work has received funding in part from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101001899) and in part from the German Federal Ministry of Education and Research (BMBF) within the project Open6GHub (grant agreement 16KISK010).