α Belief Propagation as Fully Factorized Approximation

Dong Liu*, Nima N. Moghadam†, Lars K. Rasmussen*, Jinliang Huang†, Saikat Chatterjee*

* KTH Royal Institute of Technology, Stockholm, Sweden.
† Huawei Technologies Sweden AB, Stockholm, Sweden.
E-mail: {doli, lkra, sach}@kth.se, {nima.najari.moghadam1, jinliang.huang}@huawei.com

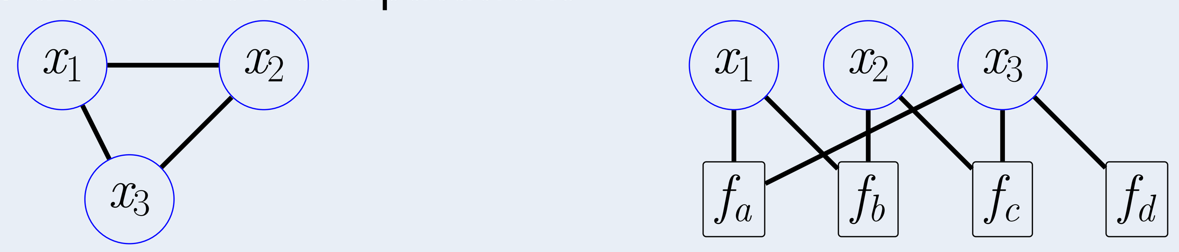
Background

Graphic Models

- Structured graphs to express conditional dependence between random variables.

Belief Propagation

- Performance inference on graphical models
- Marginal distribution computations



Two alternative ways to express dependence between random variables x_1, x_2, x_3 using graph. The dependence can be denoted by edges directly or factor nodes.

Metrics of Belief Propagation

Well known properties about belief propagation

- Exact inference on tree-structured/loop-free graphs
- Computation complexity reduction via intermediate result sharing: messages as beliefs exchange between neighboring nodes
- With proper message scheduling (loop-free), linear complexity with regarding to the size of graph

Issues remaining in standard belief propagation

- Intuition missing for graphs with loops: what is belief propagation actually doing on loopy graphs?
- Performance can degenerate significantly for graphs containing cycles

Overview of This Work

What to expect from this work?

- A new variant of belief propagation algorithm, i.e. α-BP, which generalizes standard belief propagation
- Insights of α-BP, including standard belief propagation, in general graphs
- Performance gain on cyclic graphs

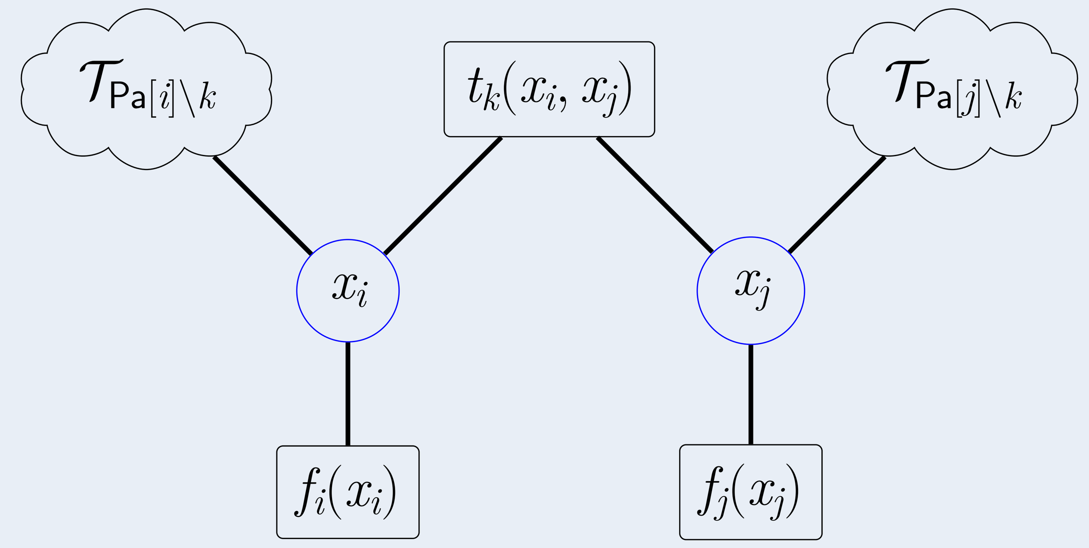
Preliminary

Pairwise Markov random field (MRF)

$$p(\boldsymbol{x}) \propto \prod_{i=1}^{N} f_i(x_i) \prod_{k \in \mathcal{K}} t_k(x_i, x_j), \boldsymbol{x} \in \mathcal{A}^N, \mathcal{A} \subset \mathbb{R}$$

- f_i is the *singleton factor*, t_k is the *pairwise factor*
- $\mathcal{K}$ is the index set of all pairwise factors

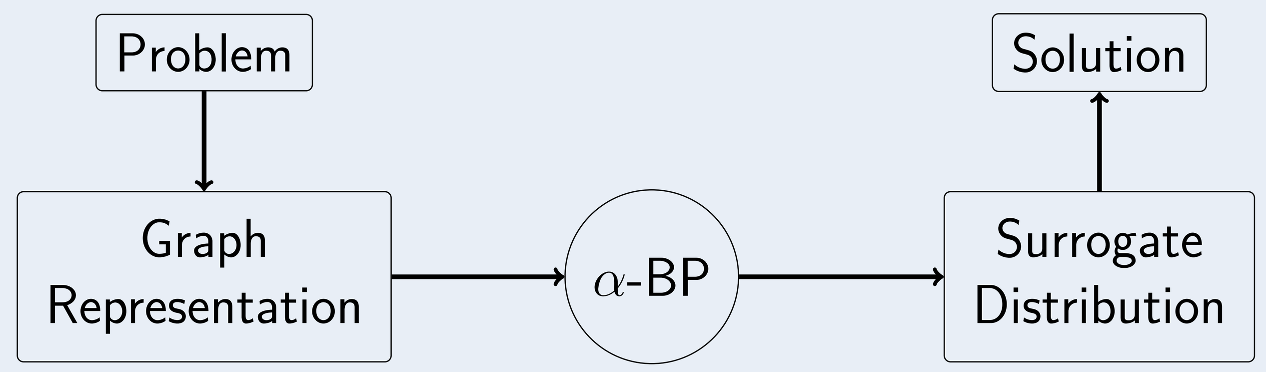
Graphical Representation



Factor graph of pairwise MRF, on which messages propagate.

- $\mathsf{Pa}[i]$ is the index set of pairwise factors connecting to variable node x_i
- $\mathcal{T}_{\mathsf{Pa}[i]\backslash k}$ is the product of all pairwise factors connecting to x_i except for t_k, $\mathcal{T}_{\mathsf{Pa}[i]\backslash k} = \prod_{n \in \mathsf{Pa}[i]\backslash k} t_n$

α-BP



- Key message rules of α-BP:

$$m_{k\to i}^{\mathsf{new}}(x_i) \propto m_{k\to i}(x_i)^{1-\alpha} \left[\sum_{x_j} t_k(x_i, x_j)^{\alpha} m_{k\to j}(x_j)^{1-\alpha} m_{j\to k}(x_j) \right]$$

where

$$m_{j\to k}(x_j) = \tilde{f}_j(x_j) \prod_{n \in \mathsf{Pa}[j]\backslash k} m_{n\to j}(x_j), \; \tilde{f}_i^{\mathsf{new}}(x_i) \propto f_i(x_i)^{\alpha} \cdot \tilde{f}_i(x_i)^{1-\alpha},$$

- What does α-BP do: message passing in graph
- Essence of α-BP: minimization of α-divergence:

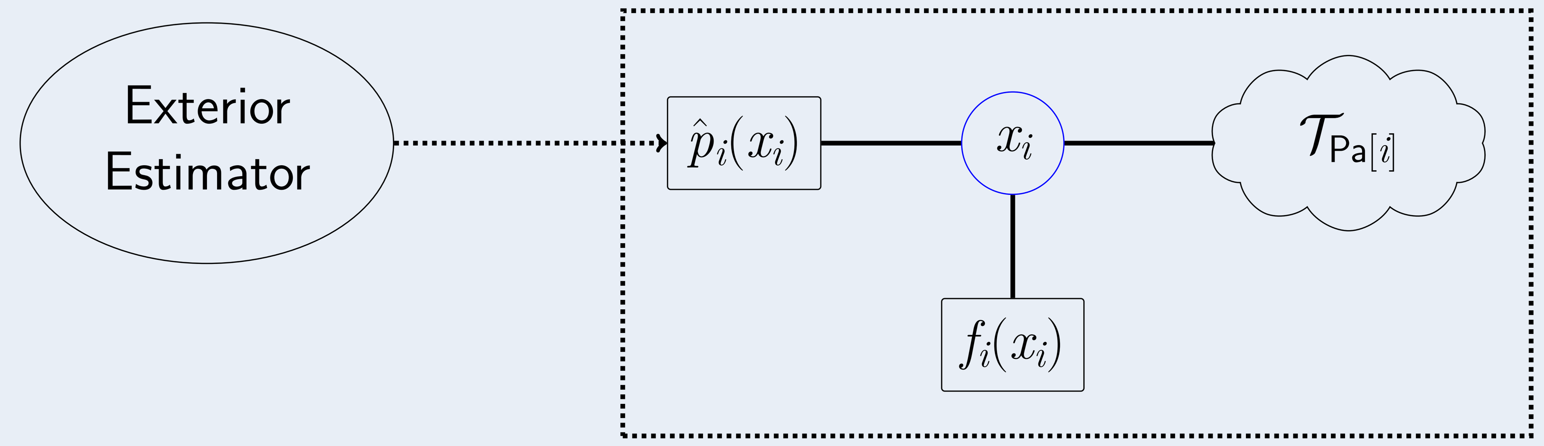
$$\mathcal{D}_{\alpha}(p\|q) = \frac{\int_{\boldsymbol{x}} \alpha p(\boldsymbol{x}) + (1-\alpha) q(\boldsymbol{x}) - p(\boldsymbol{x})^{\alpha} q(\boldsymbol{x})^{1-\alpha} d\boldsymbol{x}}{\alpha(1-\alpha)},$$

where p represents the true, complex measure of the problem, q represents the messages in graph approximating p.

α-BP Generalize BP

- α-BP reduces to BP as $\alpha \to 1$ (α-divergence reduces to Kullback-Leibler divergence).
- Minimizing Kullback-Leibler divergence leads to BP.

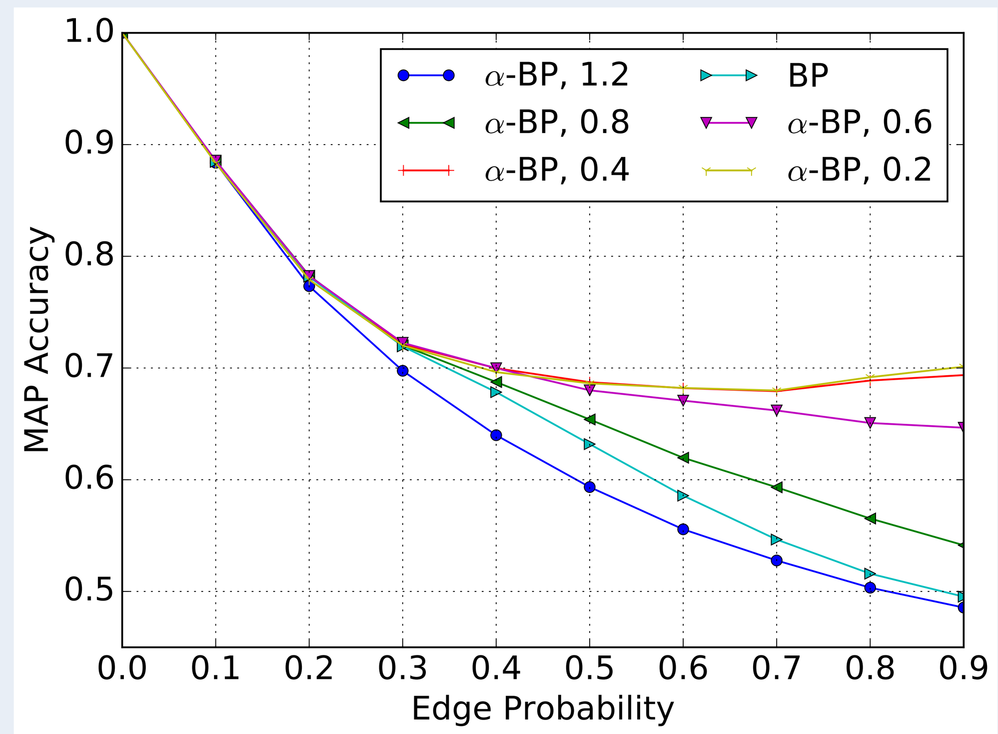
Use α-BP with Other Method



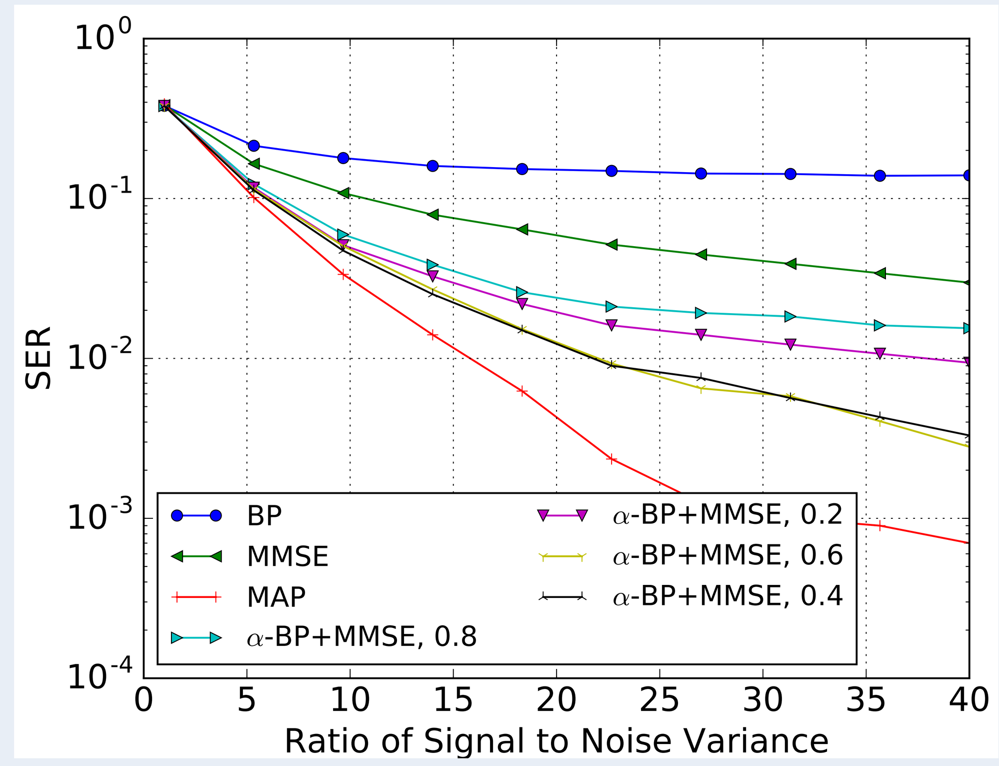
Soft combination of α-BP and other estimators.

- a) exterior estimation as prior
- b) apply α-BP on a revised graph to include prior information.

Numerical Results



Mismatch between MAP and α-BP on binary MRF



MIMO detection: α-BP with prior via MMSE

Reference

- T. P. Minka, "Expectation propagation for approximate bayesian inference," in Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, ser. UAI '01. San Francisco, CA, USA, 2001, pp.