Marginal Inference in Continuous Markov Random Fields using Mixtures

Yuanzhen Guo

Hao Xiong

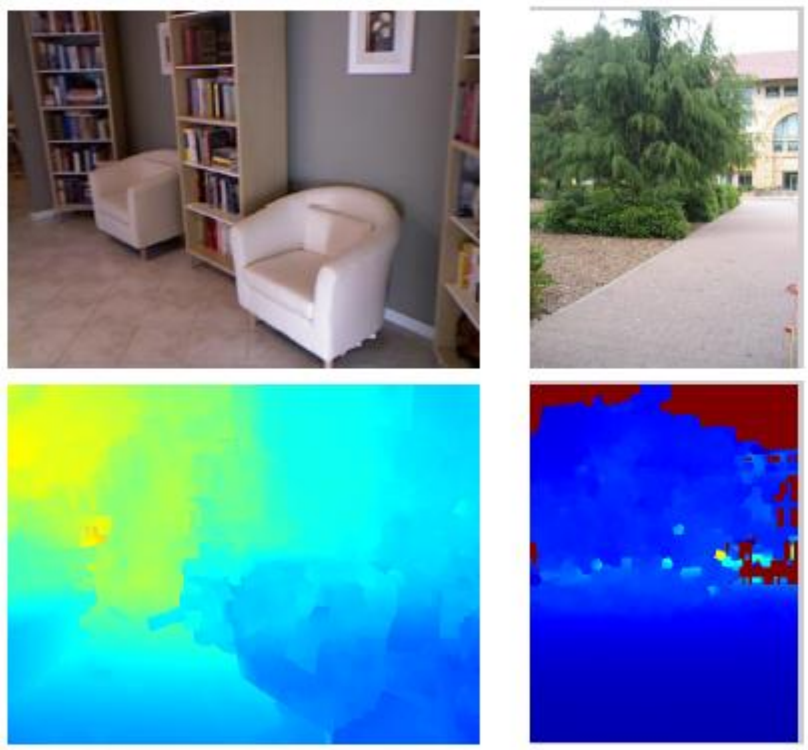
Nicholas Ruozzi

UT DALLAS
The University of Texas at Dallas

Key notes

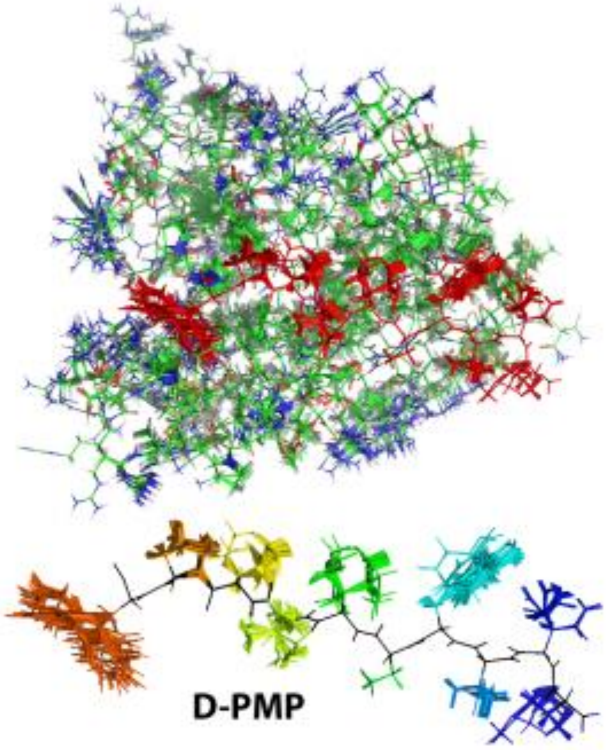
- Continuous inference in graphical models is important in scientific applications but challenging in practice
- Existing methods are extremely limited (special cases, small scales)
- Our approximate inference approach is scalable and applicable to general models
- Easy to extend to new settings with minor modifications

Continuous inference applications



Depth estimation from a single image

[Liu and etc., CVPR 2014]



Protein side chain particles and closeup of first ten amino acids

[Pacheco and Sudderth , ICML 2015]

Continuous marginal inference in MRF

- Given a graph G = (V, E) and a set of potential functions, the joint probability distribution can be written as

$$p(x_V) = \frac{1}{Z} \prod_{i \in V} \phi_i(x_i) \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)$$

- The continuous marginal inference problem are defined as

$$p(x_S) = \int_{x_V \setminus x_S} p(x_V)$$

where x_S is a subset of variables we are interested in.

Why marginal inference is hard

- Potential functions can be arbitrary (exact integration is computationally hard)
- Particle based message passing methods (PBP, EPBP, etc.)
  - Messages are not in closed forms
  - Highly depend on proposal distribution
  - Particle sampling costs too much time
- Nonparametric methods (nonparametric BP, kernel BP, etc.)
  - Scale poorly and sometimes unreasonable beliefs
- Expectation Propagation
  - Convergence/accuracy issues with simple exponential families

Our approach

- The fixed point messages of BP correspond to local optima of a constrained optimization problem known as the Bethe free energy
- Approximate beliefs directly instead of messages
- Assume beliefs over set S are mixtures of fully factorized distributions (in this case, Gaussians)

$$b_i^k(x_i) = N(x_i; \mu_{ik}, \sigma_{ik})$$

$$b_S(x_S) = \sum_k w_k \prod_{i \in S} b_i^k(x_i)$$

- The mixtures can be arbitrarily expressive
- Local marginalization constraints in BFE are naturally satisfied

Our approach

- Quadrature methods to approximate continuous integrals

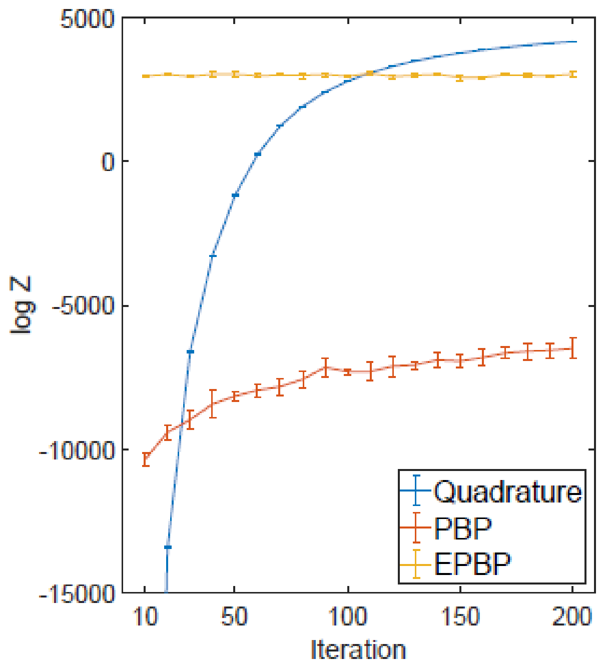
$$\int_x N(x;\mu,\sigma) f(x) \approx \sum_t w_t f(y_t)$$

- No requirement on the function f and the approximation is exact whenever f is a polynomial of degree at most $2K-1$
- Gradient ascent on Bethe Free Energy with tractable beliefs
- Marginal inference (even under evidence) is trivial
- Any marginals can be approximated by the beliefs with enough mixture components

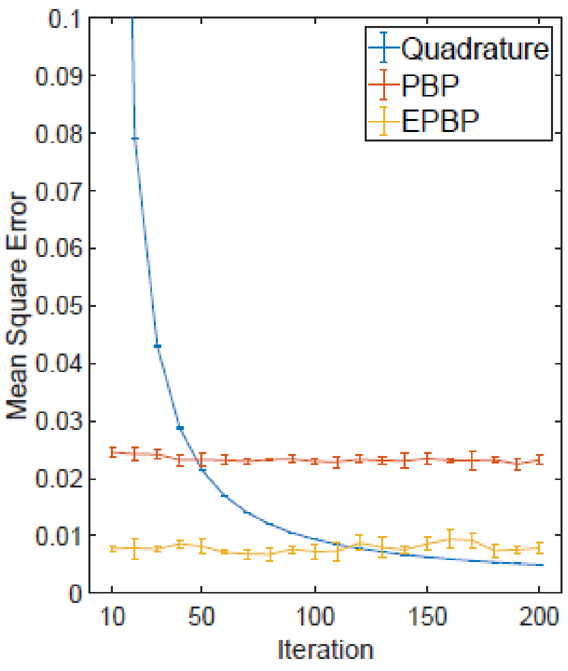
UCI datasets with Chow-Liu Trees

	Average Z			Average Univariate KL divergence		
Dataset	**PBP**	**EPBP**	**QBethe**	**PBP**	**EPBP**	**QBethe**
Iris	0.20 ± 0.17	0.37 ± 0.16	0.97 ± 0.02	0.35 ± 0.33	0.25 ± 0.14	0.00 ± 0.00
B.N.	0.15 ± 0.18	0.00 ± 0.00	0.87 ± 0.01	0.62 ± 0.59	0.83 ± 0.01	0.06 ± 0.00
I.S.E.	0.00 ± 0.01	0.06 ± 0.00	0.54 ± 0.02	0.78 ± 0.37	0.30 ± 0.00	0.21 ± 0.05
Seeds	0.12 ± 0.12	0.49 ± 0.15	0.84 ± 0.05	0.29 ± 0.18	0.12 ± 0.03	0.02 ± 0.01
Yeast	0.04 ± 0.12	0.00 ± 0.00	0.67 ± 0.07	3.31 ± 3.61	1.18 ± 0.09	0.24 ± 0.05
Wdbc	0.05 ± 0.18	0.27 ± 0.20	0.21 ± 0.06	0.10 ± 0.07	0.58 ± 0.14	0.18 ± 0.19
Letter	0.00 ± 0.00	0.00 ± 0.00	0.26 ± 0.05	0.57 ± 0.26	0.73 ± 0.01	0.07 ± 0.02
Poker	0.62 ± 0.12	0.01 ± 0.00	0.63 ± 0.05	0.02 ± 0.01	0.32 ± 0.00	0.06 ± 0.01
CMSC	0.32 ± 0.08	0.47 ± 0.01	0.56 ± 0.02	0.03 ± 0.01	0.02 ± 0.00	0.02 ± 0.00

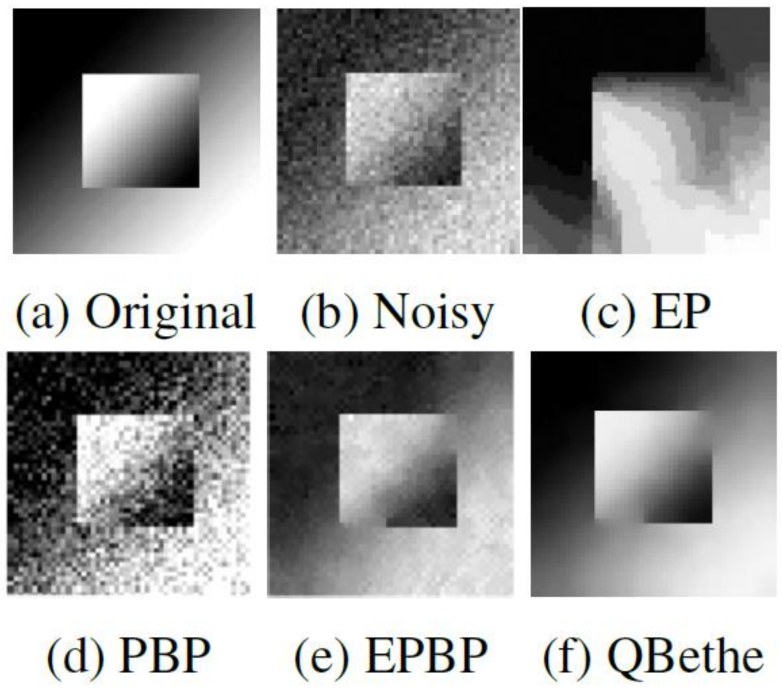
Image denoising



Log-partition function



Mean square error



Denoised images

Scalability

- Optical flow model on 584x388 grid graph with over 200,000 random variables
- With a GPU implementation, our method completed 5,000 iterations in just under 200 seconds (.04 sec/iteration)

Original Image QBethe Ground truth

Beyond marginals

- With fully factorized beliefs, MAP and Marginal MAP problems can be easily handled
- Unlike message passing algorithms, our method is able to capture global information
- Naturally extend to continuous-discrete mixed case as well as can be easily vectorized
- From inference to learning with maximize likelihood

Thank you