Probabilistic Circuits for Variational Inference in Discrete Graphical Models

Andy Shih, Stefano Ermon
{andyshih,ermon}@cs.stanford.edu

NEURAL INFORMATION PROCESSING SYSTEMS

Variational Inference

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Any choice of q gives a lower bound

Choice of q

Analytic optimization:

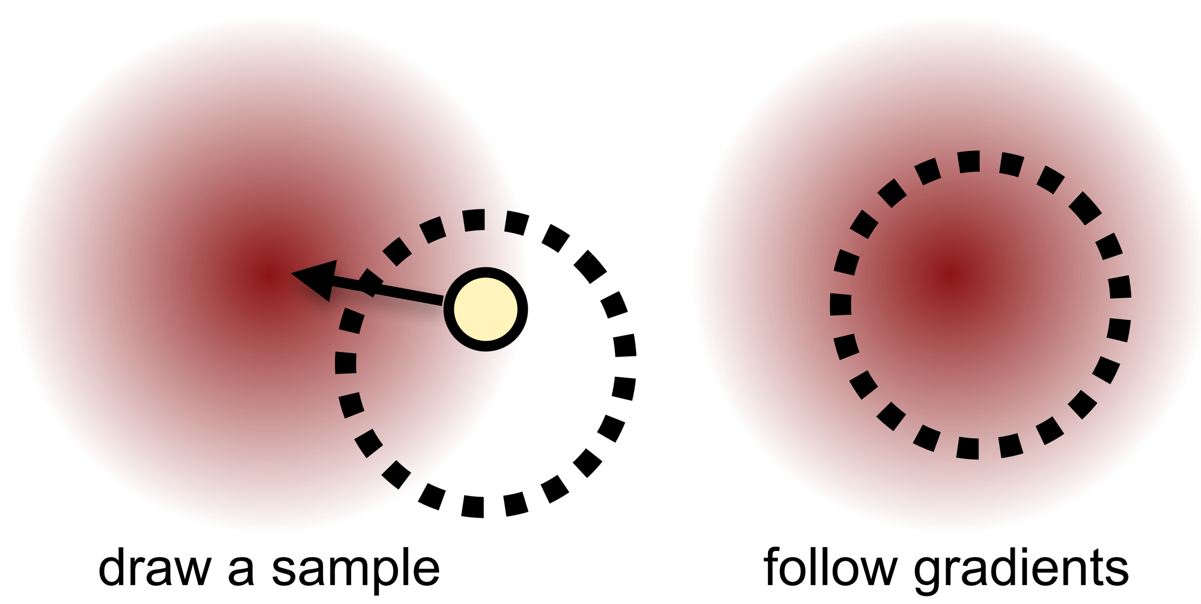
- mean field / structured mean field

Stochastic optimization:

- neural networks
- more expressive, but requires *sampling*

✓ continuous ✗ discrete [Zhang 2017]

Continuous Settings

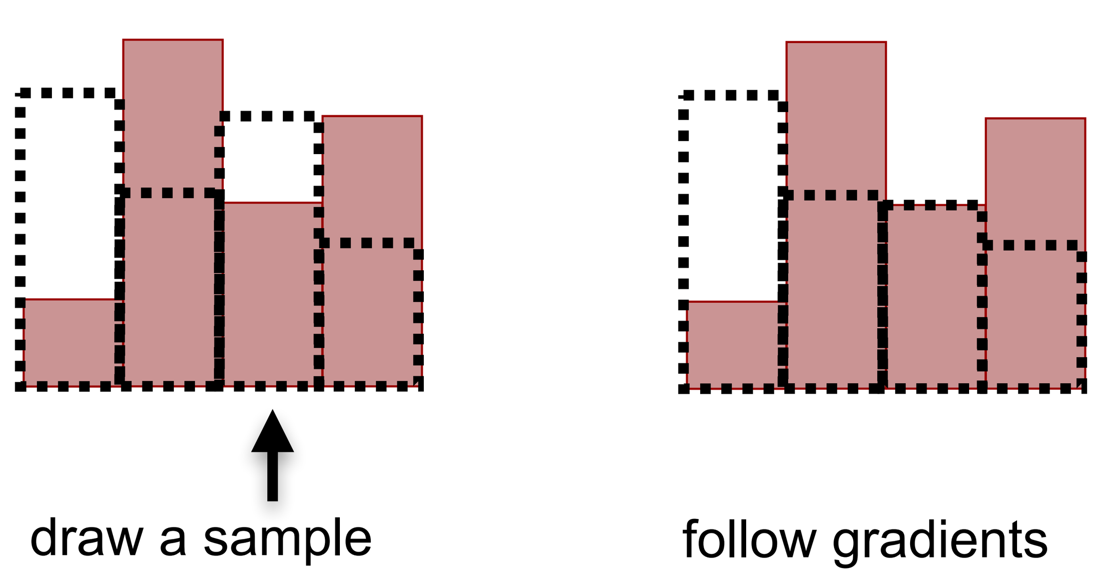


Sampling works well

Reparameterization trick: low variance

✓

Discrete Settings

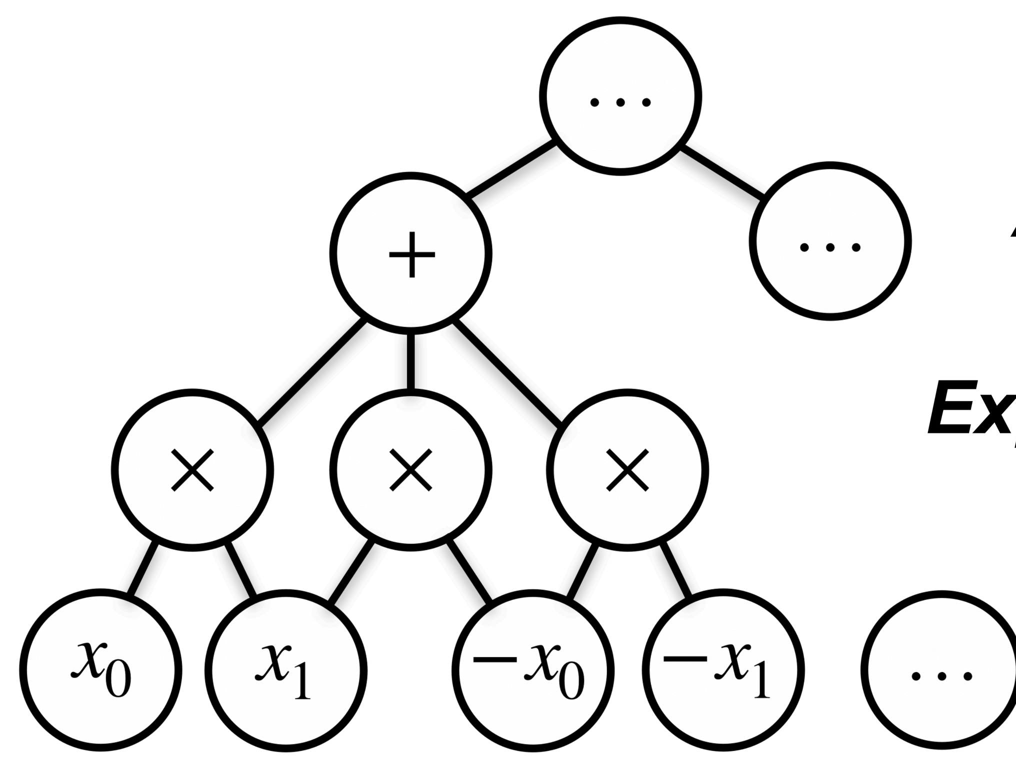


No info "around" the sample

Large variance in high dimensions

✗

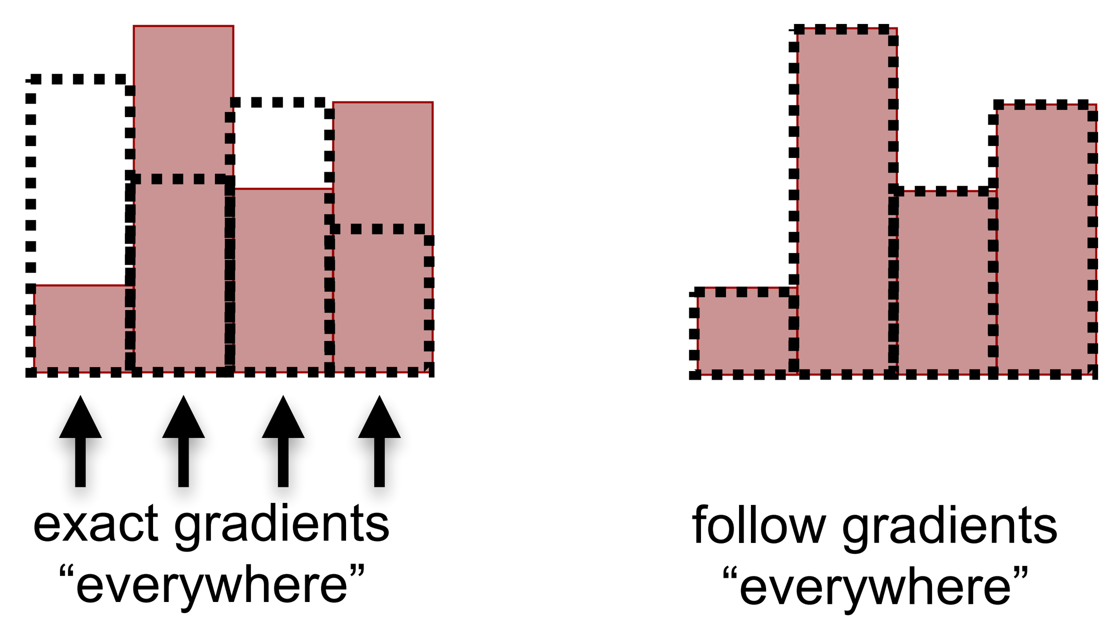
Sum Product Networks



Avoids sampling

Expressive distribution

✓ **discrete settings**



Sum Node

Selectivity / Determinism:

$$-\mathbb{E}_{\mathbf{x}\sim g_i}\left[\log g_i(\mathbf{x})\right] = -\sum_{j\in ch(i)} \alpha_{ij}\log\alpha_{ij} + \alpha_{ij}\mathbb{E}_{\mathbf{x}\sim g_j}\left[\log g_j(\mathbf{x})\right]$$

entropy of sum — *sum of (constant terms + entropy)*

Product Node

Children are independent

Computes the expectation analytically:

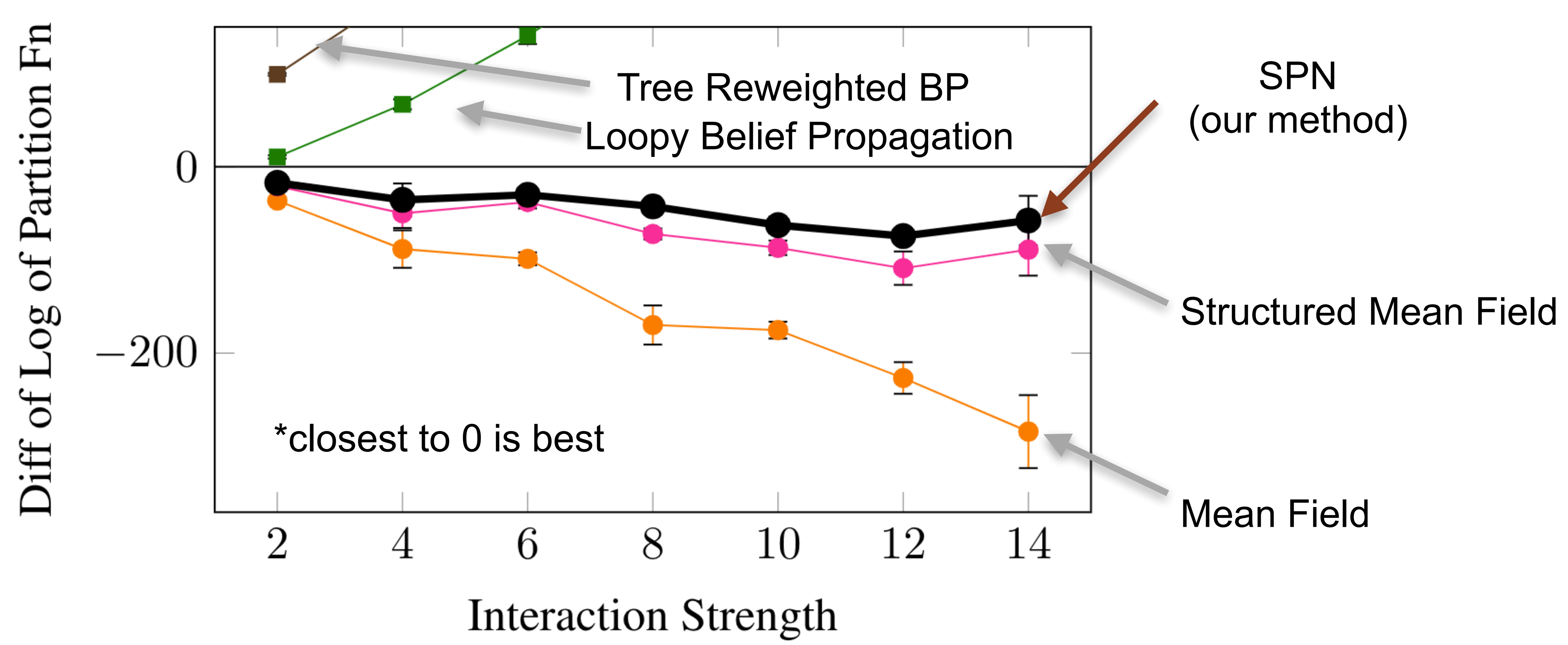
$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Computation cost: $O(tm)$ $O(m)$ ← *previously* $O(m^2)$ [Lowd 2010]

t: size of discrete model, m: size of SPN

↖ expressed in log-polynomial form

Experiments



Paper / Code