Gauging Variational Inference

Sunsoo Ahn[1], Michael Chertkov[2], Jinwoo Shin[1]

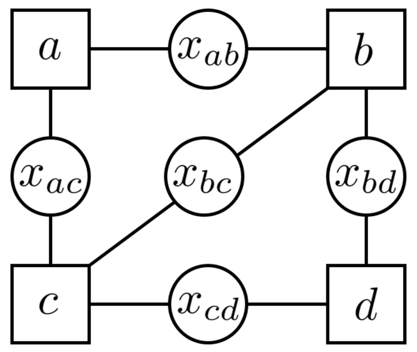
[1]Korea Advanced Institute of Science and Technology (KAIST)

[2] Los Alamos National Laboratory and Skolkovo Institute of Science and Technology

Goal: Approximating the Partition Function

Forney-style graphical model (GM) express distributions by graph $G = (V, E)$, where (binary) variable correspond to edge and factor to vertex:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{a\in V} f_a(\mathbf{x}_{\partial a}),$$
$$Z := \sum_{\mathbf{x}\in\{0,1\}^E} \prod_{a\in V} f_a(\mathbf{x}_{\partial a}),$$



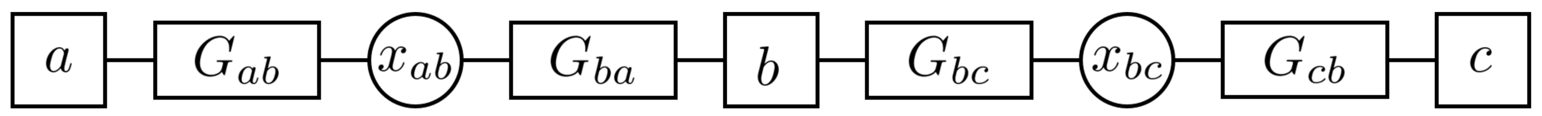
- Partition function Z is essential, but #P-hard to approximate.
- Forney-style representation is universal, i.e., any high-order GM can be expressed as Forney style.

Most popular variational algorithms for approximating Z:

- **Mean-field (MF) approach**
  Lower bounding algorithm with relatively bad approximation quality.
- **Belief propagation (BP)**
  Good approximation quality, no guarantee on bounding Z.

Gauge Transformation of Graphical Model

Gauge transformation (GT) is linear transformation of factors, leaving partition function Z invariant.



- GT is defined with respect to pairs of 2×2 matrices (G_{ab}, G_{ba}) called gauges, associated with each edge $(a, b) \in E$:

$$G_{ab} = \begin{bmatrix} G_{ab}(0,0) & G_{ab}(0,1) \\ G_{ab}(1,0) & G_{ab}(1,1) \end{bmatrix}, \quad G_{ba} = (G_{ab}^\top)^{-1}$$

- Given $\mathcal{G} = \{G_{ab}, G_{ba} : (a, b) \in E\}$, factor is transformed as follows:

$$f_a(\mathbf{x}_{\partial a}) \to f_{a,\mathcal{G}}(\mathbf{x}_{\partial a}) = \sum_{\mathbf{x}'_{\partial a}\in\{0,1\}^{\partial a}} f_a(\mathbf{x}'_{\partial a}) \prod_{b\in\partial a} G_{ab}(x_{ab}, x'_{ab}),$$

- Distribution of GM: $p(\mathbf{x}) \to p_\mathcal{G}(\mathbf{x}) = \prod_{a\in V}(\mathbf{x}_{\partial a})/Z_\mathcal{G}$ where $Z_\mathcal{G} = Z$.
- E.g., in above figure, $f_{b,\mathcal{G}} = G_{ba} f_b G_{bc}^\top$, $f_{a,\mathcal{G}} = G_{ab} f_a$, $f_{c,\mathcal{G}} = G_{cb} f_c$,

$$Z_\mathcal{G} = (f_a G_{ab}^\top)(G_{ba} f_b G_{bc}^\top)(G_{cb} f_c) = Z.$$

Our Contribution

By introducing GT as an additional degree of freedom, we propose improved version of MF and BP as follows:

- **Gauged-Mean Field** improves approximation quality of MF.
- **Gauged-Belief Propagation** corrects BP to lower bound Z.

Gauged-Mean Field (G-MF)

We propose **G-MF**, which maximize MF lower bound of $\log Z$ while searching GT for GM with tight MF lower bound:

$$\underset{q,\mathcal{G}}{\text{maximize}} \quad \sum_{\mathbf{x}\in\{0,1\}^E} q(x) \log \frac{\prod_{a\in V} f_{a,\mathcal{G}}(\mathbf{x}_{\partial a})}{q(x)}$$
$$\text{such that} \quad q(x) = \prod_{(a,b)\in E} q(x_{ab}), \quad G_{ab}^\top G_{ba} = \mathbb{I} \quad \forall (a,b) \in E.$$

- Reduce to MF when every GT is identity matrix, e.g., $G_{ab} = \mathbb{I}$.
- Becomes an uncontrained optimization by plugging $G_{ba} \leftarrow (G_{ab}^\top)^{-1}$.
- We alternatively optimized MF distributions (q) and gauges ($\mathcal{G}$) , via standard MF algorithm and generic (e.g., IPOPT) solver respectively.

Gauged-Belief Propagation (G-BP)

GT express BP approximation for $\log Z$ as stationary points of following optimization problem:

$$\underset{\mathcal{G}}{\text{maximize}} \quad \log \prod_{a\in V} f_{a,\mathcal{G}}(0, 0, \cdots),$$
$$\text{such that} \quad G_{ab}^\top G_{ba} = \mathbb{I} \quad \forall (a,b) \in E.$$

- Equivalent to maximizing a single term in GM.
- No gaurantee on lower bounding $\log Z$ since factors may be negative.

We propose **G-BP**, adding non-negativity constraints on factors:

$$f_{a,\mathcal{G}}(\mathbf{x}_{\partial a}) \geq 0 \quad \forall \mathbf{x}_{\partial a} \in \{0,1\}^{\partial a}.$$

- We add log-barrier terms corresponding to non-negative constraints, and then optimize via generic solver.

Correction schemes for G-BP

We also propose schemes for improving the G-BP lowerbound, i.e., $\prod_{a\in V} f_{a,\mathcal{G}}(0, 0, \cdots) \leq Z$, as follows:

- **G-BP-single**: Counting additional, multiple 'correction' terms:

$$\prod_{a\in V} f_{a,\mathcal{G}}(0, 0, \cdots) + \sum_{i=1}^{|E|} \prod_{a\in V} f_{a,\mathcal{G}}(\mathbf{x}^{(i)}_{\partial a}),$$

  where $\mathbf{x}^{(i)} = [x_{e_i} = 1, x_{e_j} = 0, \forall j \neq i]$.
- **G-BP-sequential**: Considering $\binom{|E|}{2}$ correction terms, i.e., consider both $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(i_1,i_2)} = [x_{e_{i_1}} = 1, x_{e_{i_2}} = 1, x_{e_j} = 0, \forall j \neq i_1, i_2]$.
- **G-BP-sequential**: Conditioning to obtain auxiliary GM with Z':

$$\prod_{a\in V} f_{a,\mathcal{G}}(0, 0, \cdots) + Z',$$

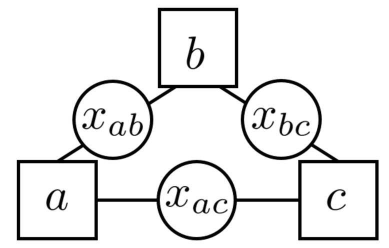
  where G-BP is applied to $Z' = \sum_{\mathbf{x}:x_{e_i}=1} \prod_{a\in V} f_{a,\mathcal{G}}(\mathbf{x}_{\partial a})$ again.

Optimality of G-MF and G-BP

Theorem. (Ahn, Chertkov, Shin 2017) *Gauged-MF and Gauged-BP formulation outputs the exact partition function Z for:*

1. *GMs defined on any line graph.*
2. *alternating cycle GM (where MF, BP perform bad).*

- Example of alternating cycle GM:



$$f_a = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, f_b = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, f_c = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

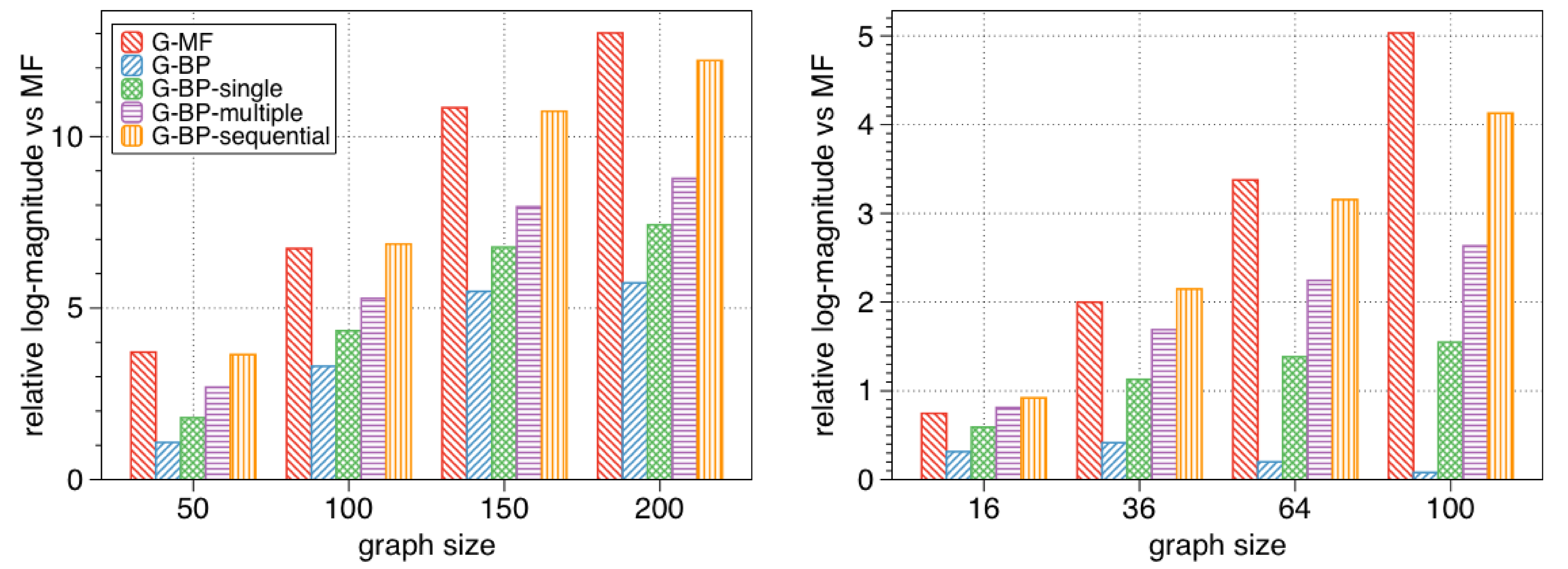
Experiments

- We consider G-MF, G-BP, and corrected versions of G-BP (i.e., G-BP-single, G-BP-multiple, G-BP-sequential).
- Factors are prepared by 'interaction strength' parameters $\{\beta_a\}_{a\in V}$:

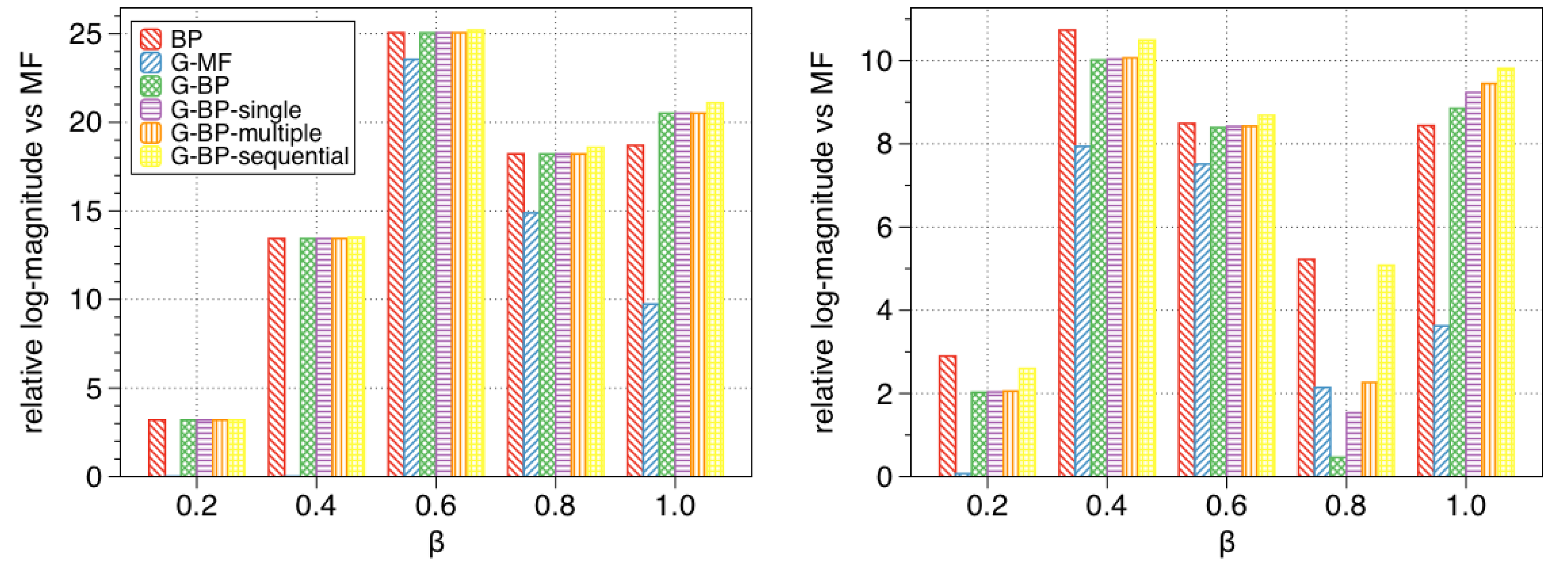
$$f_a(\mathbf{x}_{\partial a}) = \exp(\beta_a|(\text{\# of ‘0’s in } \mathbf{x}_{\partial a}) - (\text{\# of ‘1’s in } \mathbf{x}_{\partial a})|).$$

Experiment results on 3-regular (left) and grid(right) graphs:

- With varying graph size $|V|$:



- On log-supermodular factors with varying strengths, i.e., $\beta > 0$:



Conclusion

We propose two gauge optimizations:

- **Gauged-MF**, improving approximation quality of MF.
- **Gauged-BP**, modifying BP to provide lower bounds of Z.

Our results have large potential for generalization.