Gauged Mini-Bucket Elimination for Approximate Inference

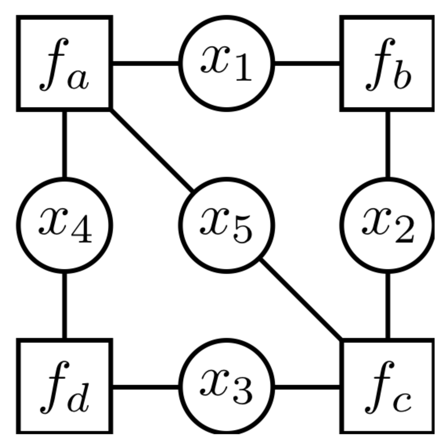
Sunsoo Ahn[1], Michael Chertkov[2], Jinwoo Shin[1], Adrian Weller[3]

[1]Korea Advanced Institute of Science and Technology
[2]Los Alamos National Laboratory, Skolkovo Institute of Science and Technology
[3]University of Cambridge, The Alan Turing Institute

Goal: Approximating the Partition Function

Graphical model (GM) is associated with factor graph $G=(V,E)$ with vertices $V = X \cup F$ with discrete variables X and factors F.

$$p(\mathbf{x}) = \frac{1}{Z}\prod_{\alpha\in F} f_\alpha(\mathbf{x}_\alpha),$$
$$Z := \sum_{\mathbf{x}\in\mathcal{X}^E}\prod_{\alpha\in F} f_\alpha(\mathbf{x}_\alpha),$$



- Partition function Z is #P-hard to compute.
- We consider Forney-style GMs, where $|N(v)| = 2$ for all $v \in X$.

Weighted Mini-Bucket Elimination

Bucket elimination (BE) computes exact Z by sequential elimination:

1. Pick variable x_v and neighboring factors (bucket) B_v.
2. Update new factor f_{B_v} as follows:
$$f_{B_v}(\mathbf{x}_{B_v}) = \sum_{x_v}\prod_{f_\alpha\in B_v} f_\alpha(\mathbf{x}_\alpha),$$
however size of f_{B_v} may grow exponentially large.

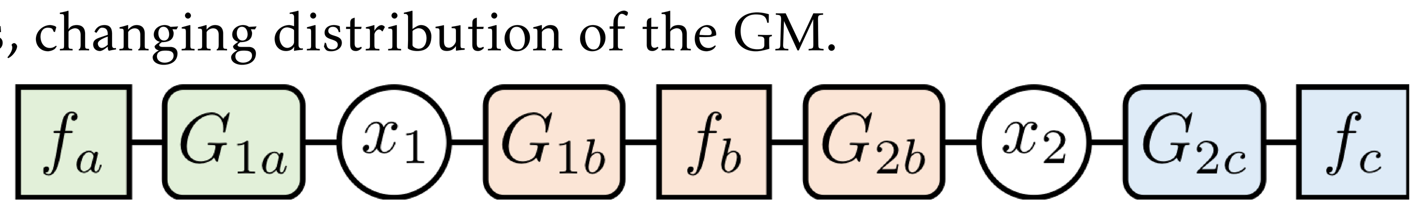
Weighted mini-bucket elimination (WMBE) approximates BE:

2⋆. If $|B_v| > ibound$, update new factors $\{f_{B_v^r}\}_{r=1}^{R_v}$ as follows:
$$f_{B_v^r}(\mathbf{x}_{B_v^r}) = \left(\sum_{x_v}\prod_{f_\alpha\in B_v^r} |f_\alpha(\mathbf{x}_\alpha)|^{\frac{1}{w_r}}\right)^{w_r},$$
for $r = 1,\cdots,R_v$ where $\sum_{r=1}^{R_v} w_r = 1$, $\{B_v^r\}_{r=1}^{R_v}$ is partition of B_v.

WMBE upper-bounds Z based on *Hölder's inequality*:
$$\sum_{x_v}\prod_{f_\alpha\in B_v} f_\alpha(\mathbf{x}_\alpha) \le \prod_{r=1}^{R_v}\left(\sum_{x_v}\prod_{f_\alpha\in B_v^r}|f_\alpha(\mathbf{x}_\alpha)|^{\frac{1}{w_r}}\right)^{w_r}.$$

Gauge Transformation of Graphical Models

Gauge transformation (GT) is Z-invariant linear transformation for factors, changing distribution of the GM.



- Defined by set of matrices $\mathcal{G} = \{G_{v\alpha} : (v,\alpha)\in E\}$, termed *gauges*:
$$G_{v\alpha} = \begin{bmatrix} G_{v\alpha}(1,1) & \cdots & G_{v\alpha}(1,d) \\ \vdots & \ddots & \vdots \\ G_{v\alpha}(d,1) & \cdots & G_{v\alpha}(d,d)\end{bmatrix},$$
where $G_{v\alpha}^\top G_{v\beta} = \mathbb{I}$ for $N(v) = \{\alpha,\beta\}$.
- Each factor is transformed as follows:
$$\widehat{f}_\alpha(\mathbf{x}_\alpha;\mathbf{G}_\alpha) = \sum_{\mathbf{x}'_\alpha} f_\alpha(\mathbf{x}'_\alpha)\prod_{v\in N(\alpha)} G_{v\alpha}(x_v, x'_v).$$
- Reduce to original GM when $G_{v\alpha} = \mathbb{I}$ for all $G_{v\alpha}\in\mathcal{G}$.

Contribution

We propose **Gauge transformation (GT)** framework for improving the accuracy of WMBE algorithm.

- outperforms existing variational schemes for WMBE.
- generalizes the **reparameterization** framework.

Gauged Weighted Mini-Bucket Elimination

We propose **Gauged WMBE** algorithm, which minimize the upper bound of WMBE Z_{WMBE}:

$$\underset{\mathcal{G}}{\text{maximize}} \quad Z_{\text{WMBE}}(\mathcal{G})$$
$$\text{such that} \quad G_{v\alpha}^\top G_{v\beta} = \mathbb{I} \quad \forall v\in X, N(v) = \{\alpha,\beta\}.$$

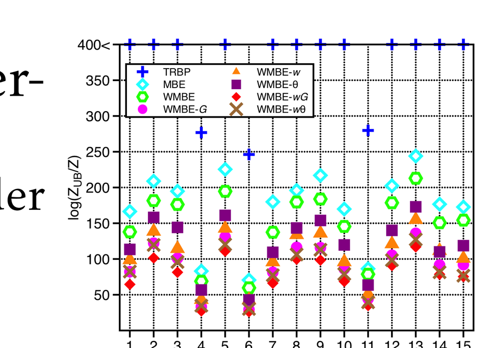
- Becomes unconstrained by plugging $G_{v\beta} \leftarrow (G_{v\alpha}^{-1})^\top$.
- Gradient descent for optimization:
  1. Initialize gauges via $G_{v\alpha} \leftarrow \mathbb{I}$.
  2. Update Gauge gradients for all $G_{v\alpha}$ via message passing:
  $$G_{v\alpha}(x'_v, x''_v) \leftarrow G_{v\alpha}(x'_v, x''_v) - \mu\frac{\partial \log Z_{\text{WMBE}}(\mathcal{G})}{\partial G_{v\alpha}(x'_v, x''_v)},$$
  3. Gauge-transform GM, i.e., $f_\alpha(\mathbf{x}_\alpha) \leftarrow \widehat{f}(\mathbf{x};\mathbf{G}_\alpha)$. Go to step 1.

Theoretical Results

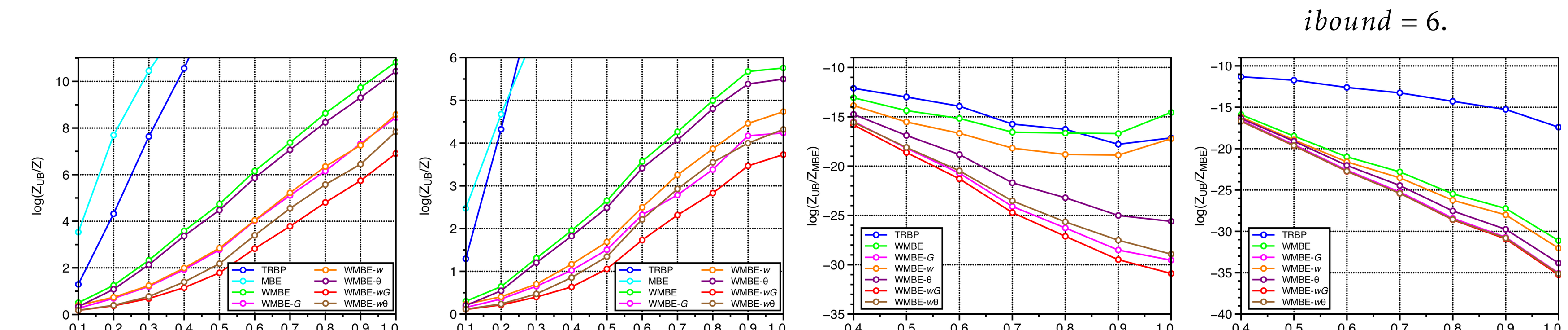
1. GT is generalization for **reparameterization** of GMs, defined on set of vectors $\boldsymbol{\theta} = \{\theta_{v\alpha} : (v,\alpha)\in E\}$:
$$\widehat{f}_\alpha(\mathbf{x}_\alpha;\boldsymbol{\theta}_\alpha) = \prod_{v\in N(\alpha)}\exp(\theta_{v\alpha}(x_i)) f_\alpha(\mathbf{x}_\alpha),$$
where $\theta_{v\alpha} + \theta_{v\beta} = 0$ for $N(v) = \alpha,\beta$.
2. Unlike GT, reparameterization fails in binary symmetric GMs, where distribution is invariant to 'flipping' of variables.

Experimental Results

1. We compare our algorithm with other upper-bounding algorithms for Z:
   - WMBE optimized with Gauge (WMBE-G), Hölder weight (WMBE-w), reparameterization (WMBE-θ)
   - Tree reweighted belief propagation (TRBP)
   - Mini-bucket elimination (MBE)

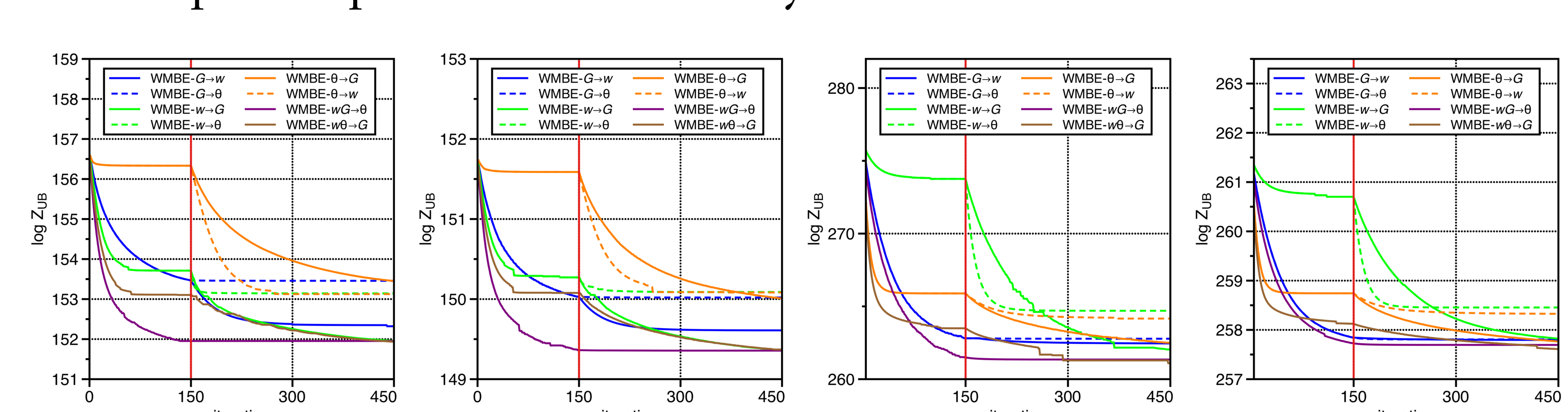


UAI 2014 Linkage dataset *ibound* = 6.



Ising grid GMs, *ibound* = 4 | Ising grid GMs, *ibound* = 6 | 3-regular GMs, *ibound* = 4 | 3-regular GMs, *ibound* = 6

2. To measure effectiveness of optimizing over parameters, we optimize each pair of parameters in one by one.



Ising grid GMs, *ibound* = 4 | Ising grid GMs, *ibound* = 6 | 3-regular GMs, *ibound* = 4 | 3-regular GMs, *ibound* = 6

Conclusion

- We developed a new scalable gauge-variational approach improving the bound of WMBE algorithm.
- Generalization to non-Forney style or continuous models would be interesting.