Bucket Renormalization for Approximate Inference

Sungsoo Ahn[1], Michael Chertkov[2,3], Adrian Weller[4,5] and Jinwoo Shin[1,6]

[1]Korea Advanced Institute of Science and Technology (KAIST), [2]Los Alamos National Laboratory (LANL), [3]Skolkovo Institute of Science and Technology (Skoltech), [4]University of Cambridge, [5]The Alan Turing Institute, [6]AITrics

ICML

SUMMARY

We consider discrete **graphical models (GMs)**, factorizing distributions by hyper-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

$$\Pr(\mathbf{x}) = \frac{1}{Z}\prod_{\alpha\in\mathcal{E}} f_\alpha(\mathbf{x}_\alpha), \quad \text{where} \quad Z = \sum_{\mathbf{x}\in\mathcal{X}^\mathcal{V}}\prod_{\alpha\in\mathcal{E}} f_\alpha(\mathbf{x}_\alpha),$$

where **partition function** Z is essential for inference, but require approximation algorithms:

- Markov chain Monte Carlo (MCMC): running Markov chains for samples.
- Variational inference (VI): casting computation of Z as an optimization.
- Approximate elimination: approximately summing out variables one-by-one.

Our contribution: Inspired from tensor renormalization group (Levin & Nave, 2007), we propose **bucket renormalization**, an approximate elimination framework outperforming precedessors and faster than VI.

BACKGROUND: APPROXIMATE ELMINIATION

Bucket (or variable) elimination (BE): For each vertex $i \in \mathcal{V}$ with bucket $\mathcal{B}_i = \{f_\alpha : i \in \alpha, \alpha \in \mathcal{E}\}$, sum out variable x_i as follows

$$f_{\mathcal{B}_i\setminus\{i\}}(\mathbf{x}_{\mathcal{B}_i\setminus\{i\}}) = \sum_{x_i\in\mathcal{X}}\prod_{f_\alpha\in\mathcal{B}_i} f_\alpha(\mathbf{x}_\alpha),$$

then $f_{\mathcal{B}_i}$ is newly added to the GM, i.e., distributive law for computing Z.

- Intractable when size of bucket $|\mathcal{V}(B_i)|$ gets large (i.e., tree-width).

Mini-bucket elimination (MBE, Decther & Rish, 2003): Replace marginalization process of BE with partitioned mini-buckets $\{\mathcal{B}_i^\ell\}_{\ell=1}^{m_i}$:

$$f_{\mathcal{B}_i^\ell\setminus\{i\}}(\mathbf{x}_{\mathcal{B}_i^\ell\setminus\{i\}}) = \max_{x_i}\prod_{f_\alpha\in\mathcal{B}_i^\ell} f_\alpha(\mathbf{x}_\alpha), \quad f_{\mathcal{B}_i^{m_i}\setminus\{i\}}(\mathbf{x}_{\mathcal{B}_i^{m_i}\setminus\{i\}}) = \sum_{x_i\in\mathcal{X}}\prod_{f_\alpha\in\mathcal{B}_i^{m_i}} f_\alpha(\mathbf{x}_\alpha),$$

for $\ell = 1, \cdots, m_i - 1$ and $\{f_{\mathcal{B}_i\setminus\{i\}^\ell}\}_{\ell=1}^{m_i}$ are newly added to the GM.

- MBE upper bounds Z since $f_{\mathcal{B}_i\setminus\{i\}} \le \prod_{\ell=1}^{m_i}\widetilde{f}_{\mathcal{B}_i^\ell\setminus\{i\}}$ for each step.
- Complexity of MBE is determined by size of mini-buckets, bounded by $|\mathcal{V}(\mathcal{B}_i^\ell)| \le \textit{ibound} + 1$, where *ibound* is called induced tree-width bound.

CONTRIBUTION: BUCKET RENORMALIZATION

Similar to MBE, we propose two algorithms that replace the marginalization process of BE with its approximations:

- **Mini-bucket renormalization (MBR)**: repeatedly renormalizing mini-buckets via rank-1 approximation of mini-buckets.
- **Global-bucket renormalization (GBR)**: calibrating the result of MBR based on explicit approximation error given by global-bucket.

Both algorithms perform superior to its predecessors, and still much faster than traditional variational inference algorithms.

MINI-BUCKET RENORMALIZATION (MBR)

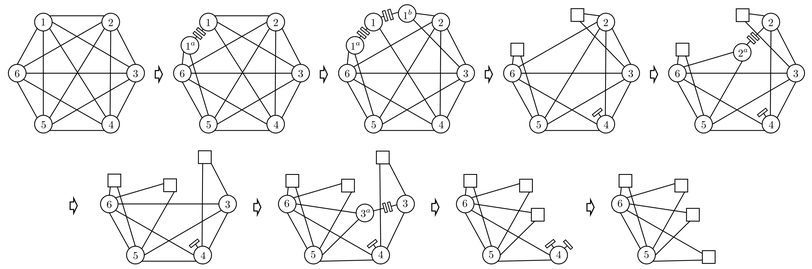


Fig: MBR with elimination order $o = 1, 2, 3, 4, 5$ and *ibound* $= 2$.

Mini-Bucket Renormalization (MBR) eliminates each vertex $i \in \mathcal{V}$ in elimination order o, with induced tree-width bound *ibound* as follows:

1. Partition the bucket $\mathcal{B}_i = \{f_\alpha : i \in \alpha, \alpha \in \mathcal{E}\}$ into mini-buckets $\{\mathcal{B}_i^\ell\}_{\ell=1}^{m_i}$ where $\mathcal{B}_i = \bigcup_{\ell=1}^{m_i}\mathcal{B}_i^\ell$, $\mathcal{B}_i^{\ell_1}\cap\mathcal{B}_i^{\ell_2} = \emptyset$ and $|\mathcal{V}(\mathcal{B}_i^\ell)| \le \textit{ibound} + 1$.

2. For $\ell = 1, \cdots, m_i - 1$, mini-bucket $\mathcal{B}_i^\ell$ is renormalized by replacing i by replicates i^ℓ and adding singleton factors r_i^ℓ, r_{i^ℓ} for error compensation:

$$\widetilde{\mathcal{B}}_i^\ell \leftarrow \{f_{\alpha\setminus\{i\}\cup\{i^\ell\}} | f_\alpha \in \mathcal{B}_i^\ell\} \cup \{r_i^\ell, r_{i^\ell}\}.$$

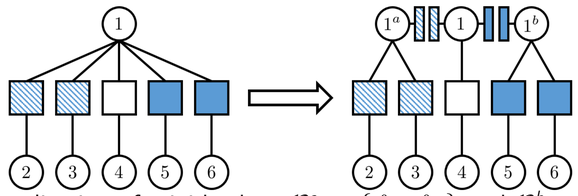


Fig: Renormalization of mini-buckets $\mathcal{B}_1^a = \{f_{12}, f_{13}\}$ and $\mathcal{B}_1^b = \{f_{15}, f_{16}\}$ into $\widetilde{\mathcal{B}}_1^a = \{f_{1^a2}, f_{1^a3}, r_1^a, r_{1^a}\}$ and $\widetilde{\mathcal{B}}_2^b = \{f_{1^b5}, f_{1^b6}, r_1^b, r_{1^b}\}$.

3. The singleton factors r_i^ℓ and r_{i^ℓ} are chosen by comparing the factor induced on $\mathbf{x}_{\mathcal{B}_i}$ from mini-bucket $\mathcal{B}_i^\ell$ and its renormalization $\widetilde{\mathcal{B}}_i^\ell$:

$$\min_{r_i^\ell, r_{i^\ell}}\sum_{\mathbf{x}_{\mathcal{B}_i^\ell}\in\mathcal{X}^{\mathcal{V}(\mathcal{B}_i^\ell)}}\left(\prod_{f_\alpha\in\mathcal{B}_i^\ell} f_\alpha(\mathbf{x}_\alpha) - \sum_{x_{i^\ell}}\prod_{f_\alpha\in\widetilde{\mathcal{B}}_i^\ell} f_\alpha(\mathbf{x}_\alpha)\right)^2,$$

which is solved via rank-1 truncated singular value decomposition (SVD) in $O(|\mathcal{X}|^{\textit{ibound}+2})$ complexity, but typically faster in existing SVD solvers.

4. Variables $x_i, x_{i^1}, \cdots, x_{i^{m_i-1}}$ are summed out seperately:

$$f_{\mathcal{B}_i^\ell\setminus\{i\}}(\mathbf{x}_{\mathcal{B}_i^\ell\setminus\{i\}}) = \sum_{x_{i^\ell}\in\mathcal{X}} r_i^\ell(x_{i^\ell})\prod_{f_\alpha\in\mathcal{B}_i^\ell} f_\alpha(x_{i^\ell}, \mathbf{x}_{\alpha\setminus\{i\}\cup\{i^\ell\}}),$$

$$f_{\mathcal{B}_i^{m_i}\setminus\{i\}}(\mathbf{x}_{\mathcal{B}_i^{m_i}\setminus\{i\}}) = \sum_{x_i\in\mathcal{X}}\prod_{\ell=1}^{m_i-1} r_i^\ell(x_{i^\ell})\prod_{f_\alpha\in\mathcal{B}_i^{m_i}} f_\alpha(\mathbf{x}_\alpha),$$

for $\ell = 1, \cdots, m_i - 1$, where computation and memory are bounded by $O(|\mathcal{X}|^{\textit{ibound}+1})$ and $\{f_{\mathcal{B}_i^\ell\setminus\{i\}}\}_{\ell=1}^{m_i}$ are newly added to the GM.

GLOBAL-BUCKET RENORMALIZATION (GBR)

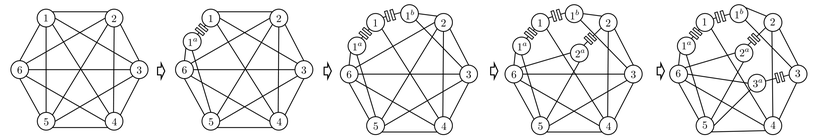


Fig: Semantics of MBR without variable marginalization

- Approximation error made at each t-th steo of MBR can be explicitly described as a GM renormalization process:

$$|Z_t(r^{(1)}, \cdots, r^{(t-1)}) - Z_{t-1}(r^{(1)}, \cdots, r^{(t)})|, \tag{1}$$

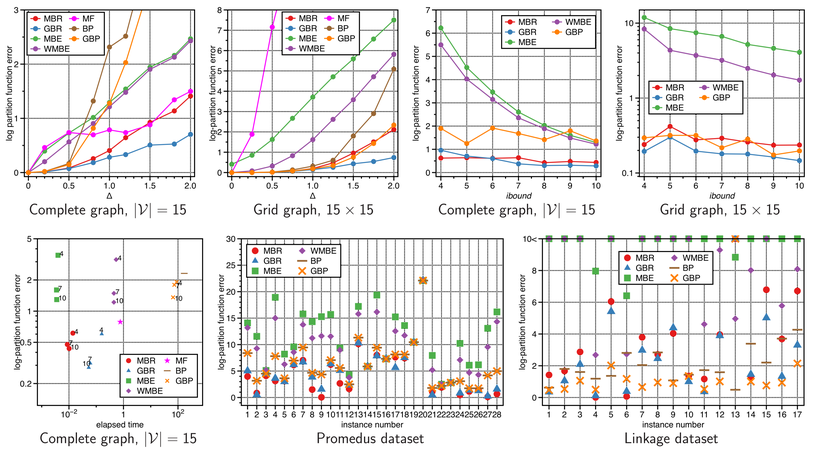
where $r^{(1)}, \cdots, r^{(t)}$ indicates the choice of singleton factors and Z_t corresponds to the partition function of t-th renormalized GM, i.e., global-bucket, with error compensating factors $r^{(t)}$.

- **Global-Bucket Renormalization (GBR)** calibrates each MBR choice of singleton factors $r^{(t)}$ by minimizing approximated value of (1), with $Z_t(\cdot)$ being approximated via applying MBR:

$$|\widetilde{Z}_t(s^{(1)}, \cdots, s^{(t-1)}) - \widetilde{Z}_{t-1}(s^{(1)}, \cdots, s^{(t)})|,$$

where $s^{(1)}, \cdots, s^{(t)}$ indicates the GBR's choice of singleton factors and $\widetilde{Z}_t \approx Z_t$ results from MBR.

EXPERIMENTS



Complete graph, $|\mathcal{V}| = 15$ | Grid graph, 15×15 | Complete graph, $|\mathcal{V}| = 15$ | Grid graph, 15×15

Complete graph, $|\mathcal{V}| = 15$ | Promedus dataset | Linkage dataset

Fig: Performance comparison to popular approximate inference algorithms.

We compare with 5 algorithms:

1. Approximate eliminations: MBE, weighted MBE (WMBE).
2. VI: belief propagation (BP), mean-field (MF), generalized BP (GBP).

Comparison was done on 2 types of GMs:

1. Ising GM (complete, grid graphs) with interaction strength Δ.
2. UAI 2014 Inference Competition datasets, named Promedus (medical diagnosis) and Linkage (genetic linkage).