Uprooting and Rerooting Higher-Order Graphical Models

UNIVERSITY OF CAMBRIDGE

Mark Rowland and Adrian Weller

The Alan Turing Institute

SUMMARY

- Rerooting maps an initial model to any of a whole family of related models.
- Inference on any one model in the family yields inference results for all models in the family.
- We provide an efficient heuristic to select a good rerooting for easier inference, yielding good empirical results.
- Theoretically, we analyse rerooting of Sherali-Adams polytopes and show that the triplet ploytope is unique in being universally rooted.

INFERENCE IN GRAPHICAL MODELS

A binary graphical model is specified by a hypergraph $G = (V, E)$, together with for each hyperedge $\mathcal{E} \in E$, a potential $\theta_\mathcal{E} : \{0,1\}^\mathcal{E} \to \mathbb{R}$. This induces a probability distribution for a set of binary random variables $(X_v)_{v \in V}$. For any $x_V \in \{0,1\}^V$:

$$p(x_V) = \frac{1}{Z}\exp\left(\text{score}(x_V)\right) , \qquad \text{score}(x_V) = \sum_{\mathcal{E} \in E} \theta_\mathcal{E}(x_\mathcal{E}) .$$

MAP inference is the problem of finding a most likely configuration:

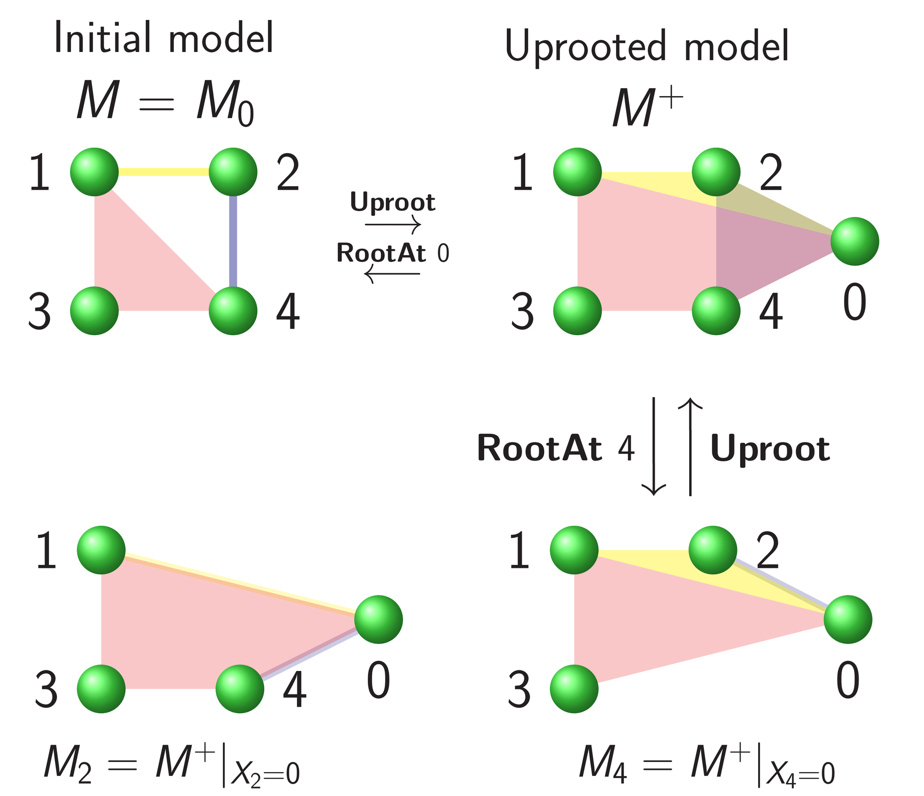
$$\arg\max_{x_V \in \{0,1\}^V} p(x_V) \left[= \arg\max_{x_V \in \{0,1\}^V} \sum_{\mathcal{E} \in E} \theta_\mathcal{E}(x_\mathcal{E}) \right] .$$

Marginal inference is the problem of computing the marginal distribution of a small collection of variables X_J, or the partition function:

$$p(x_J) = \sum_{x_{V \setminus J} \in \{0,1\}^{V \setminus J}} p(x_V) , \qquad Z = \sum_{x_V \in \{0,1\}^V} p(x_V) .$$

Both are NP-hard, motivating fast approximate inference algorithms.

UPROOTING AND REROOTING EXAMPLE



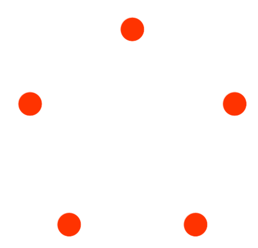
Details for pink potential shown on left

M config			M^+ configuration				M_4 config		
x_1	x_3	x_4	x_0	x_1	x_3	x_4	x_0	x_1	x_3
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	1			
0	1	0	0	0	1	0	0	0	1
0	1	1	0	0	1	1			
1	0	0	0	1	0	0	0	1	0
1	0	1	0	1	0	1			
1	1	0	0	1	1	0	0	1	1
1	1	1	0	1	1	1			
			1	0	0	0	1	0	0
			1	0	0	1			
			1	0	1	0	1	0	1
			1	0	1	1			
			1	1	0	0	1	1	0
			1	1	0	1			
			1	1	1	0	1	1	1
			1	1	1	1			

LP RELAXATIONS AND SHERALI-ADAMS POLYTOPES

Combinatorial problem of MAP inference

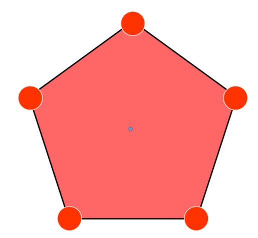
$$\max_{x \in \{0,1\}^V} \left[\sum_{\mathcal{E} \in E} \theta_\mathcal{E}(x_\mathcal{E})\right]$$



Equivalent linear program

$$\max_{(\mu_\mathcal{E} | \mathcal{E} \in E) \in \mathbb{M}(G)} \left[\sum_{\mathcal{E} \in E} \mathbb{E}_{X_\mathcal{E} \sim \mu_\mathcal{E}} \left[\theta_\mathcal{E}(X_\mathcal{E})\right]\right]$$

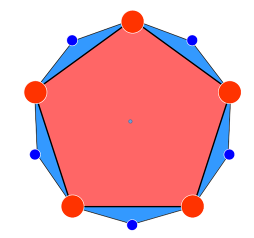
Marginal polytope $\mathbb{M}(G)$: enforces global consistency on marginals $(\mu_\mathcal{E} | \mathcal{E} \in E)$



Relaxed linear program

$$\max_{(\mu_\mathcal{E} | \mathcal{E} \in E) \in \mathbb{L}_r(G)} \left[\sum_{\mathcal{E} \in E} \mathbb{E}_{X_\mathcal{E} \sim \mu_\mathcal{E}} \left[\theta_\mathcal{E}(X_\mathcal{E})\right]\right]$$

Sherali-Adams polytope $\mathbb{L}_r(G)$: enforces consistency over each cluster of r variables on (pseudo)marginals $(\mu_\mathcal{E} | \mathcal{E} \in E)$



$\mathbb{L}_2(G)$ is the local (pairwise) polytope, and $\mathbb{L}_3(G)$ as the triplet polytope.

This yields a polytope relaxation $\mathbb{M}(G) \subseteq \mathbb{L}_r(G)$, and we therefore have

$$\max_{(\mu_\mathcal{E} | \mathcal{E} \in E) \in \mathbb{M}(G)} \left[\sum_{\mathcal{E} \in E} \mathbb{E}_{X_\mathcal{E} \sim \mu_\mathcal{E}} \left[\theta_\mathcal{E}(X_\mathcal{E})\right]\right] \leq \max_{(\mu_\mathcal{E} | \mathcal{E} \in E) \in \mathbb{L}_r(G)} \left[\sum_{\mathcal{E} \in E} \mathbb{E}_{X_\mathcal{E} \sim \mu_\mathcal{E}} \left[\theta_\mathcal{E}(X_\mathcal{E})\right]\right]$$

TRI IS UNIQUE IN BEING UNIVERSALLY ROOTED

Weller (2016) observed that $\mathbb{L}_3$ is universally rooted, in that LP+$\mathbb{L}_3$ yields the same optimum score on all rerootings.

We introduce symmetrised Sherali-Adams polytopes for uprooted models: $\widetilde{\mathbb{L}}^0_{r+1}(G)$ enforces consistency on all $r+1$ clusters that include the vertex 0.

Uprooting and rerooting induces a correspondence between polytopes:

$$\begin{array}{llllll} \text{For } M = M_0 : & \mathbb{L}_{r+1}(G) & \subseteq & \text{Unnamed} & \subseteq & \mathbb{L}_r(G) \\ & \text{Uproot} \downarrow \uparrow \text{RootAt } 0 & & \text{Uproot} \downarrow \uparrow \text{RootAt } 0 & & \text{Uproot} \downarrow \uparrow \text{RootAt } 0 \\ \text{For } M^+ : & \widetilde{\mathbb{L}}^0_{r+2}(\nabla G) & \subseteq & \widetilde{\mathbb{L}}_{r+1}(\nabla G) & \subseteq & \widetilde{\mathbb{L}}^0_{r+1}(\nabla G) . \end{array}$$

In this language, we provide a strengthening of Weller's earlier result:

> **Theorem** ("$\mathbb{L}_3$ is universally rooted")
>
> $\widetilde{\mathbb{L}}^0_4(\nabla G) = \widetilde{\mathbb{L}}_3(\nabla G)$ for all graphs G.

It was natural to conjecture that this property might hold for all Sherali-Adams polytopes of degree ≥ 3. Perhaps surprisingly, we show that:

> **Theorem**
>
> $\mathbb{L}_3(G)$ is **unique** in being universally rooted.

EXPERIMENTAL RESULTS

We estimate the partition function Z via the HAK method (Heskes et al., 2003) using libDAI (Mooij, 2010).

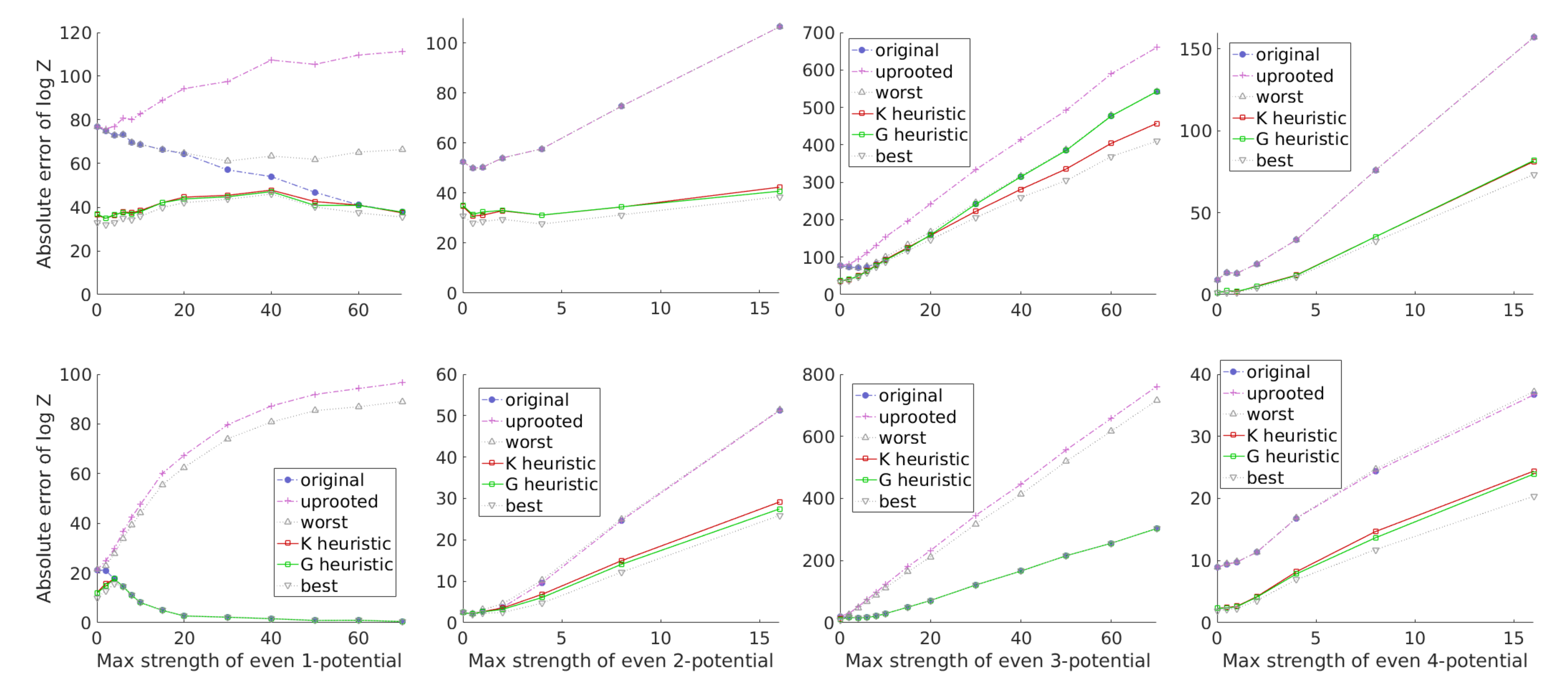
We compare results of approximate inference on the original model, uprooted model, best and worst rerooted models, and rerooted models chosen according to a heuristic of the form:

$$\arg\max_{i \in V} \left[\sum_{i \in \mathcal{E} : |\mathcal{E}| = 2} c_2 \tanh |t_2 a_\mathcal{E}| + \sum_{i \in \mathcal{E} : |\mathcal{E}| = 4} c_4 \tanh |t_4 a_\mathcal{E}| \right] .$$

The heuristic aims to identify when rerooting will be beneficial, and which variable to reroot at if so – empirically it works well.

The constants c_2, c_4, t_2, t_4 are estimated via an evolutionary algorithm. The terms $a_\mathcal{E}$ are strengths of "pure" cluster interactions.

The plots below show error in recovering $\log Z$ for complete graphs (top row) and grid graphs (bottom row), whilst varying strengths of 1–, 2–, 3–, and 4– pure potentials:



REFERENCES

T. Heskes, K. Albers, and B. Kappen. Approximate inference and constrained optimization. In *UAI*, pages 313–320, 2003.

J. Mooij. libDAI: A free and open source C++ library for discrete approximate inference in graphical models. *Journal of Machine Learning Research*, 11:2169–2173, August 2010. URL `http://www.jmlr.org/papers/volume11/mooij10a/mooij10a.pdf`.

A. Weller. Uprooting and rerooting graphical models. In *International Conference on Machine Learning (ICML)*, 2016.

ACKNOWLEDGEMENTS

MR acknowledges support by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L016516/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis. AW acknowledges support by the Alan Turing Institute under the EPSRC grant EP/N510129/1, and by the Leverhulme Trust via the CFI.