ETH zürich

Learning & Adaptive Systems

Variable Clamping for Optimization-Based Inference

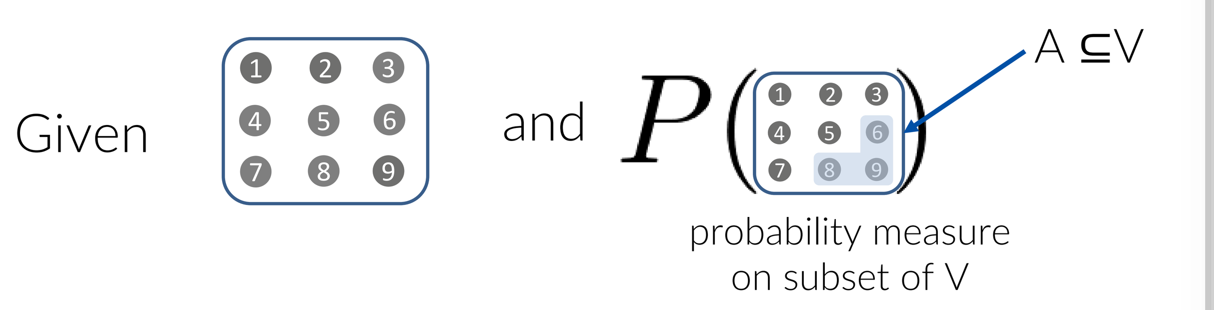
Junyao Zhao, Josip Djolonga, Sebastian Tschiatschek, Andreas Krause

Contributions

- We investigate the improvements obtained by variable clamping for two approximate inference techniques.
- We theoretically prove that the following bounds on the partition function can be only improved by clamping
  - L-Field for multi-label log-supermodular models
  - Perturb-and-MAP for general binary models
- We propose a set of heuristic strategies for selecting clamping variables and empirically showcase the improvements obtained by the proposed methods.

The Main Idea

Setup



- We assume a distribution of the form

$$P(A) = \frac{\exp(-f(A))}{\mathcal{Z}(f)},\ A \subseteq V$$

with the partition function

$$\mathcal{Z}(f) = \sum_{A \subseteq V} e^{-f(A)}$$

- #P-hard to do inference in such generic models

MAP

- Assume: efficient minimization of $f(A) + z(A)$

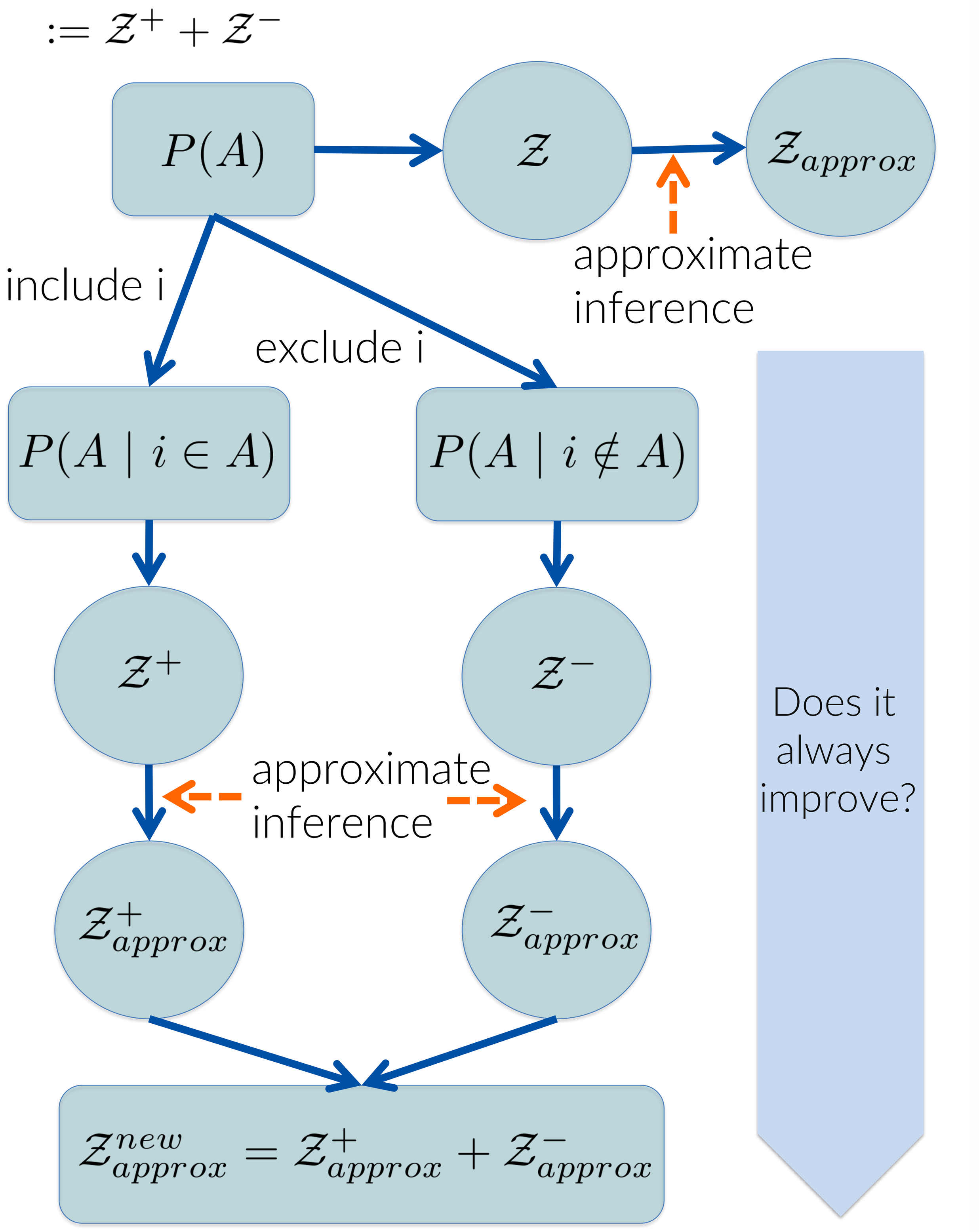
($z(A) = \sum_{i \in A} z_i$ is an arbitrary modular function)

Variable Clamping

- It builds on the observation that

$$\mathcal{Z} = \sum_{A \subseteq V} e^{-f(A)} = \sum_{A \subseteq V, i \in A} e^{-f(A)} + \sum_{A \subseteq V, i \notin A} e^{-f(A)}$$

$$:= \mathcal{Z}^+ + \mathcal{Z}^-$$

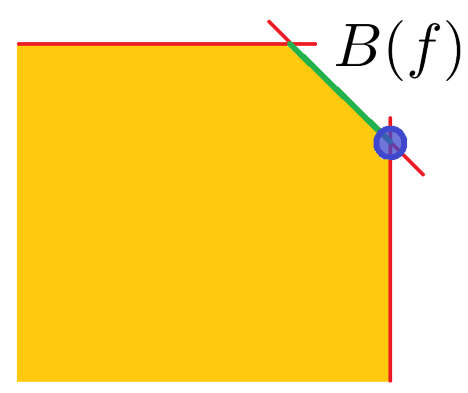


Clamping with L-Field and Perturb-and-MAP

L-Field

- Assumes log-supermodular distributions, i.e. f is submodular.

$$A, B \subseteq V \quad f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$$



- L-Field minimizes $D_\infty(P \parallel Q) = \log \sup_{S \subseteq V} \frac{P(S)}{Q(S)}$
- Or equivalently, solves

$$\log Z_L(f) = \min_{\mathbf{s} \in B(f)} \sum_{i=1}^{n} \log(1 + e^{-s_i})$$

where $B(f)$ is the base polytope, which is well understood

Perturb-and-MAP

- Only assumes efficient minimization of $f(A) + z(A)$
- Perturb-and-MAP upper bound:

$$\log Z_P(f) = \mathbb{E}_{\mathbf{z}}[\max_{A \subseteq V} z(A) - f(A)]$$

(each coordinate of $\mathbf{z}$ is sampled independently from a logistic distribution)

Main Results

Theorem. [Informal] Clamping can only decrease the upper bound of the log-partition function attained from L-Field for *multi-label log-supermodular models*.

Theorem. [Informal] Clamping can only decrease the upper bound of the log-partition function attained from Perturb-and-MAP for *general binary models*.

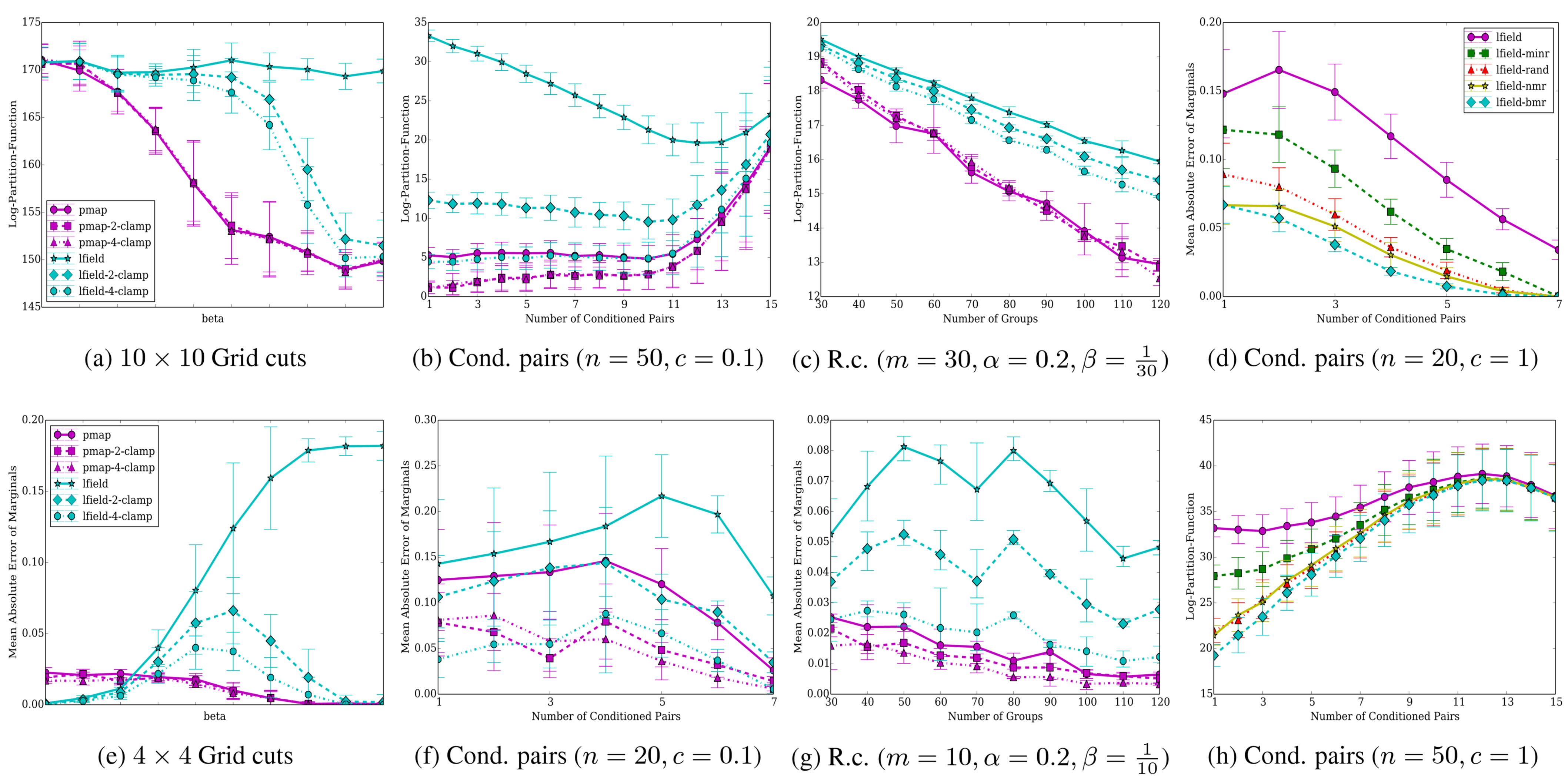
Clamping Strategies

KEY INSIGHT: Given that the bound is an optimization problem over $B(f)$, it might make sense to clamp those variables that can "vary" the most.

- Observation: $s_i \in [f(V) - f(V \setminus \{i\}), f(\{i\})],\ \forall i, \forall \mathbf{s} \in B(f)$
- Strategy: choose the variable which maximizes the size of this range

Experiments

- Three different log-supermodular models: **grid cuts**, **conditioned pairs**, **random covers**.
- We also compare different heuristics for selecting variable to clamp.



(a) 10×10 Grid cuts (b) Cond. pairs ($n = 50, c = 0.1$) (c) R.c. ($m = 30, \alpha = 0.2, \beta = \frac{1}{30}$) (d) Cond. pairs ($n = 20, c = 1$)

(e) 4×4 Grid cuts (f) Cond. pairs ($n = 20, c = 0.1$) (g) R.c. ($m = 10, \alpha = 0.2, \beta = \frac{1}{10}$) (h) Cond. pairs ($n = 50, c = 1$)

Figure 1: In the first three columns we show the effects on the estimated partition function (first row) and marginals (second row). We can see that clamping improves the estimates on both Z and the marginals. In last column we compare the proposed clamping strategies for L-Field. As evident from the plots, `bmr` consistently outperforms the other proposed alternatives.