Bethe and Related Pairwise Entropy Approximations

Adrian Weller, University of Cambridge

SUMMARY

Belief propagation may be viewed as a heuristic to optimize the Bethe free energy $\mathcal{F}_B$, and often performs strikingly well. Here we focus on binary pairwise MRFs, and generalize important results on the Bethe approximation to the broad family of related pairwise entropy approximations with arbitrary counting numbers. We make observations that shed light on the success of the Bethe approximation.

KEY RESULTS, FOR ANY COUNTING NUMBERS

All extend earlier results that were specifically for the Bethe approximation:

- Given singleton marginals, we provide an analytic solution for optimum pairwise pseudomarginals.
- Thus, the approximate free energy $\mathcal{F}_A$ may be considered a function only of singleton pseudomarginals $\{q_i = q(X_i = 1)\}$.
- We provide upper and lower bounds on first derivatives $\frac{\partial \mathcal{F}_A}{\partial q_i}$ as a function of q_i, that hold for all values of other marginals $\{q_j : j \neq i\}$, see Figure.
- We use these derivative bounds to construct an *ϵ-sufficient mesh* over pseudomarginals such that the optimum of $\mathcal{F}_A$ on the mesh is guaranteed to have value within ϵ of the global optimum, see Figures.
- We derive all second derivatives of the approximate free energy $\frac{\partial^2 \mathcal{F}_A}{\partial q_i \partial q_j}$.
- Using second derivatives, we show that for attractive models, the discrete optimization problem is *submodular*, hence may be solved efficiently leading to a FPTAS for the approximate log-partition function $\log Z_A$ (extends to balanced models).

BACKGROUND

We consider a binary pairwise model with variables $\mathcal{V}$ and edges $\mathcal{E}$.

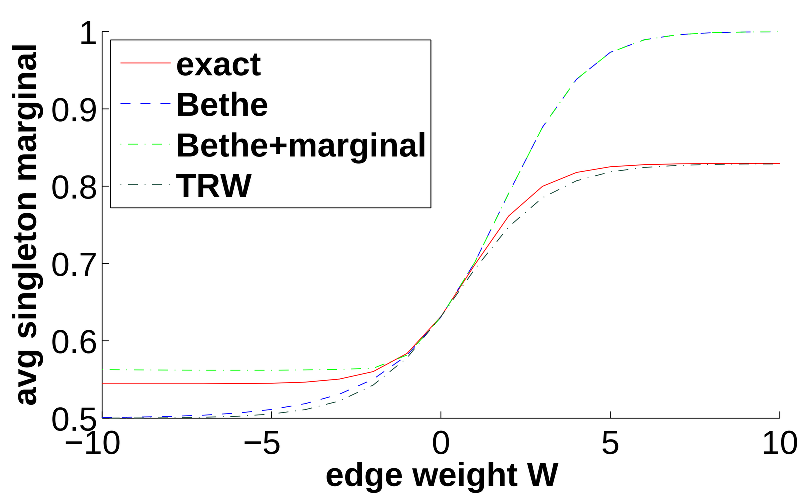
$$p(x) = \frac{e^{-E(x)}}{Z}, \qquad E = -\sum_{i\in\mathcal{V}} \theta_i x_i - \sum_{(i,j)\in\mathcal{E}} W_{ij} x_i x_j, \qquad x_i \in \{0,1\}.$$

- Variational exact inference: $-\log Z = \min_{\mu\in\mathbb{M}} \mathbb{E}_\mu(E) - S(\mu)$, S is the true entropy.
- Approximate inference: $-\log Z_A = \min_{\mu\in\mathbb{L}} \mathbb{E}_\mu(E) - S_A(\mu)$
- Approximate free energy: $\mathcal{F}_A = \mathbb{E}_\mu(E) - S_A(\mu)$
- Approximate entropy: $S_A = \sum_{i\in\mathcal{V}} c_i S_i(\mu_i) - \sum_{(i,j)\in\mathcal{E}} \rho_{ij} I_{ij}$, for arbitrary counting numbers $\{c_i, \rho_{ij} \in \mathbb{R}\}$.
- $I_{ij} = S_i(\mu_i) + S_j(\mu_j) - S_{ij}(\mu_{ij}) \geq 0$ is the mutual information.
- For example, Bethe approximation: $c_i = 1\ \forall i, \rho_{ij} = 1\ \forall (i,j)$.
- Tree-reweighted approximation TRW: $c_i = 1\ \forall i, \rho_{ij} \leq 1\ \forall (i,j)$.

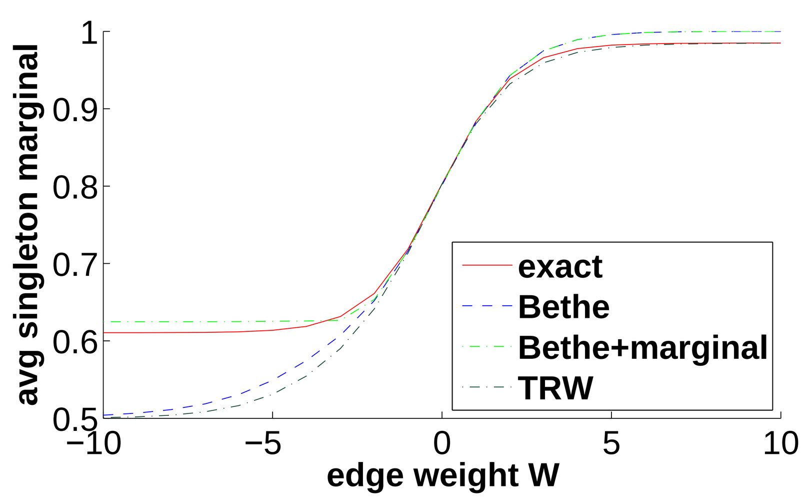
UNDERSTANDING APPROXIMATION ERROR

We focus here on the Bethe and TRW approximations. TRW has a convex free energy and leads to an upper bound on the true partition function.

1. Marginals Any non-convex free energy approximation (such as Bethe) can lead to singleton marginals pulled towards 0 or 1 as edge weights rise above a threshold. For example, a triangle with uniform edge weight W:



$\theta_i \sim U[0,1]$



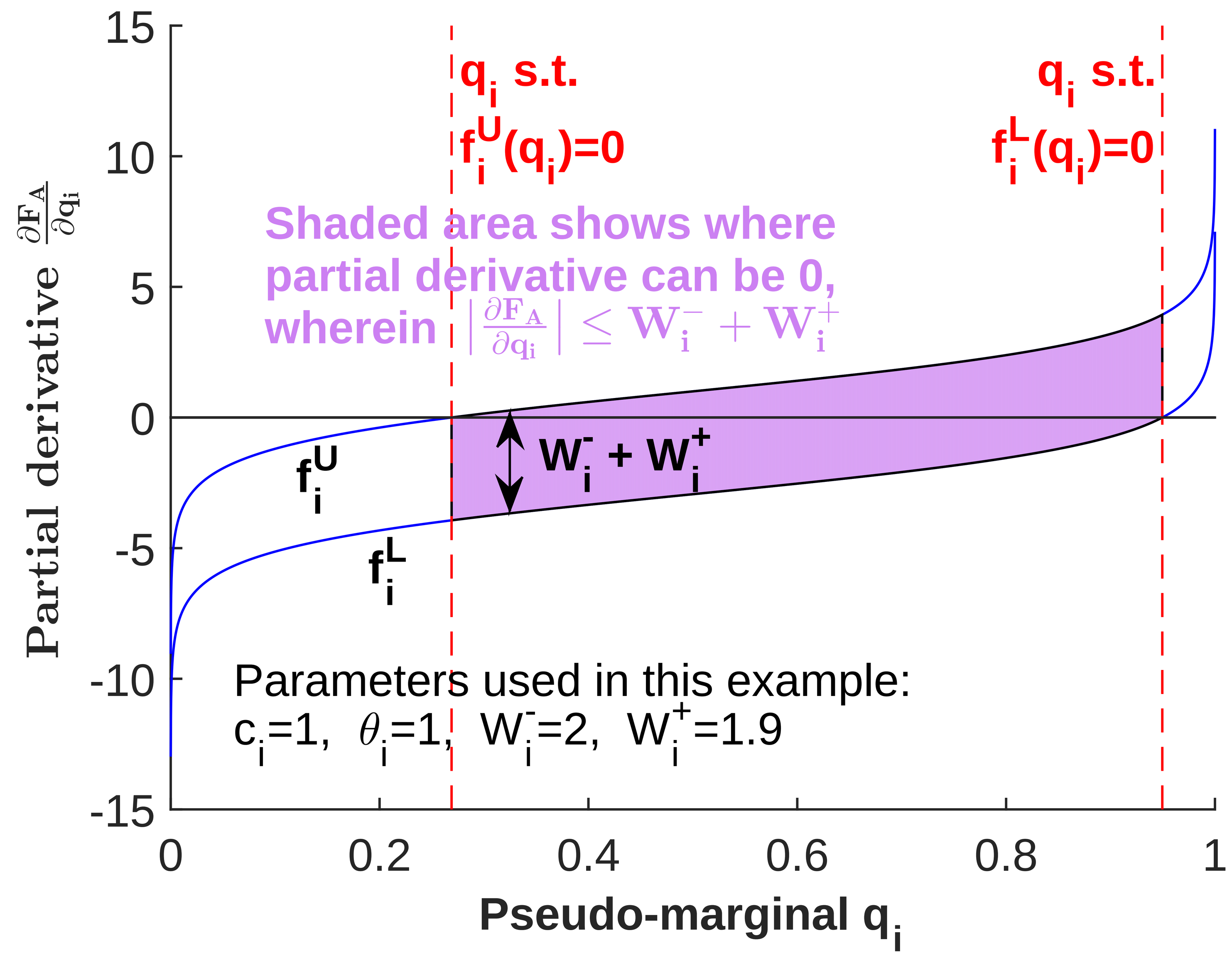
$\theta_i \sim U[0,3]$

A *polytope* effect pushes the other way, towards $\frac{1}{2}$, for *frustrated cycles*. Hence, it may be helpful to have Bethe's 'balancing' pull towards 0 or 1.

2. Partition function Consider the Bethe and TRW approximations. Both are exact for acyclic models. If there is exactly one cycle, which is balanced, it is known that $\frac{1}{2}Z \leq Z_B \leq Z$ (so Z_B is low but not by much); if the cycle is frustrated then $Z \leq Z_B$ using loop series methods, leading to high error with no upper bound.

Since $S_B \leq S_T$, $Z_B \leq Z_T$ and hence, for a frustrated cycle, $Z \leq Z_B \leq Z_T$. This is one reason for caution when estimating partition functions with approximations such as TRW.

UPPER f_i^U AND LOWER f_i^L BOUNDS FOR FIRST DERIV $\frac{\partial \mathcal{F}_A}{\partial q_i}$



Theorem *For any counting numbers, assuming optimum pairwise pseudomarginals, first derivatives of $\mathcal{F}_A$ are sandwiched in the range*

$$-\theta_i + c_i \log\frac{q_i}{1-q_i} - W_i^+ \leq \frac{\partial \mathcal{F}_A}{\partial q_i} \leq -\theta_i + c_i \log\frac{q_i}{1-q_i} + W_i^-,$$

where $W_i^+ = \sum_{j\in\mathcal{N}(i):W_{ij}\geq 0} W_{ij}$ (positive incident weights), $W_i^- = \sum_{j\in\mathcal{N}(i):W_{ij}\leq 0} -W_{ij}$.

MESH ALGORITHM TO APPROXIMATE LOG Z_A TO ANY ACCURACY FOR ANY COUNTING NUMBERS

Input: ϵ, model parameters $\{\theta_i, W_{ij}\}$ and counting numbers $\{c_i, \rho_{ij}\}$

Output: Estimate of global optimum $\log Z_A$ guaranteed in $[\log Z_A - \epsilon, \log Z_A]$, with corresponding pseudomarginals

(1) For each X_i: Compute upper and lower bound curves for $\frac{\partial \mathcal{F}_A}{\partial q_i}$, shrink search space to region where $\frac{\partial \mathcal{F}_A}{\partial q_i}$ can be 0, see Figures.

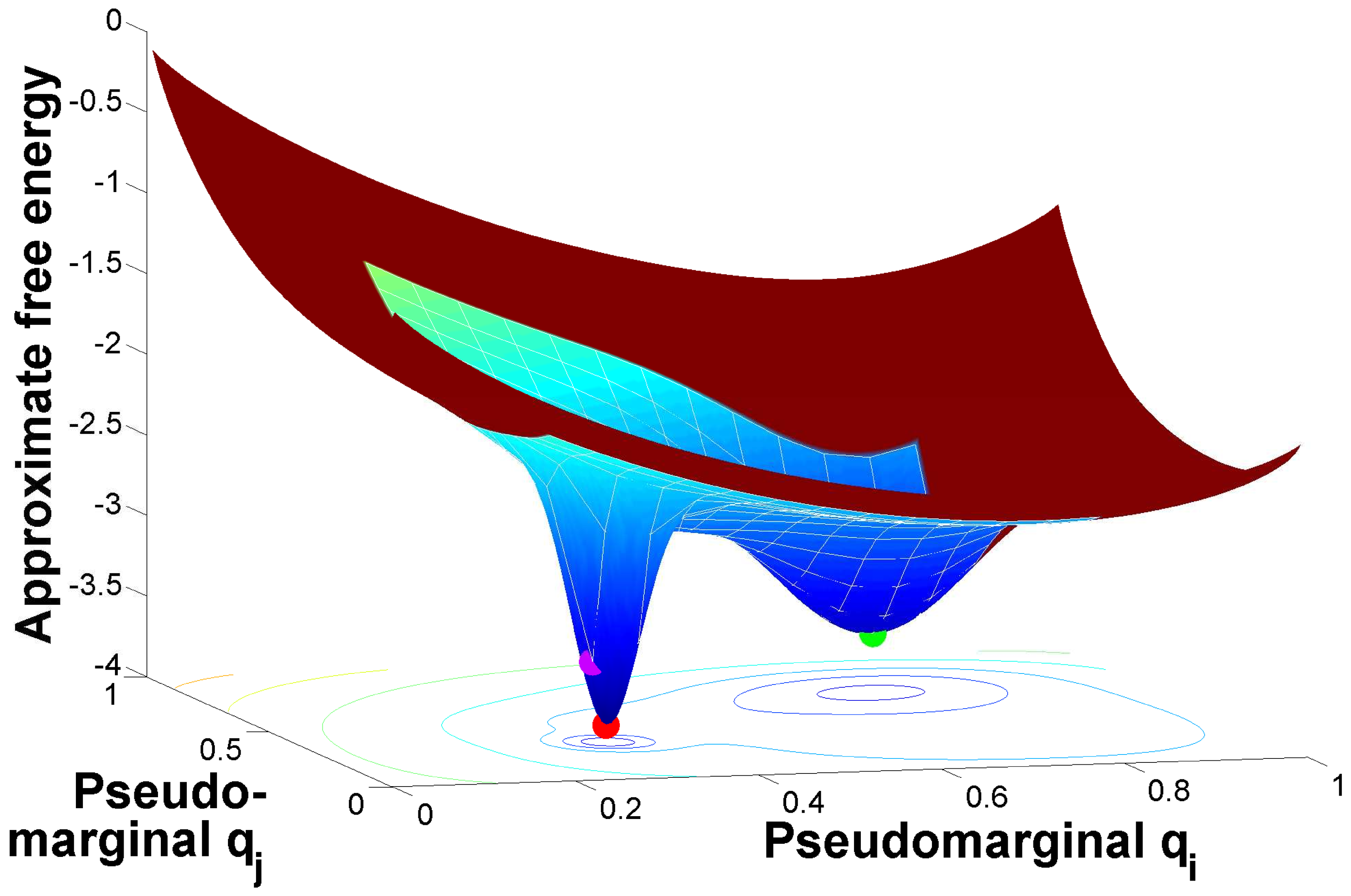
(2) Construct an ϵ-sufficient mesh using $\left|\frac{\partial \mathcal{F}_A}{\partial q_i}\right| \leq W_i^- + W_i^+$.

(3) Solve the resulting discrete optimization problem: if the model is attractive (any counting numbers) then efficient since submodular.

Theorem *For any counting numbers and any discretization, an attractive model yields a submodular discrete optimization problem to estimate $\log Z_A$.*

APPROXIMATE FREE ENERGY LANDSCAPE (STYLIZED)

We may ignore the red region and search only over the blue region with an ϵ-sufficient mesh, see above. We return the green dot (discretized optimum).



REFERENCES

E. Sudderth, M. Wainwright, and A. Willsky. Loop series and Bethe variational bounds in attractive graphical models. In *NIPS*, 2007.

A. Weller and T. Jebara. Bethe bounds and approximating the global optimum. In *Artificial Intelligence and Statistics (AISTATS)*, 2013.

A. Weller and T. Jebara. Approximating the Bethe partition function. In *UAI*, 2014a.

A. Weller and T. Jebara. Clamping variables and approximate inference. In *NIPS*, 2014b.

M. Welling and Y. Teh. Belief optimization for binary networks: A stable alternative to loopy belief propagation. In *Uncertainty in Artificial Intelligence (UAI)*, 2001.