Efficient Exact Inference in Planar Ising Models

Nicol N. Schraudolph and Dmitry Kamenetsky

Contributions

Correspondence in binary undirected graphical models
cut(*planar* Ising model) ⇔ **perfect matching**(expanded dual)
yields polynomial-time exact inference algorithms:

- **ground states** (MAP, margin violators)
  - as fast and scalable as graph cuts, but
  - need planarity instead of submodularity
- **partition function** (Globerson/Jaakkola 2007) and **marginals**
- **ML**& **MM** (max-margin) **CRF training**: grids up to 100×100

We provide unified **framework**, improved **algorithms**, & **code**:
http://aps.arxiv.org/abs/0810.4401 **http://nic.schraudolph.org/isinf/**

Model

Generic Model

- Undirected graph $G(\mathcal{V},\mathcal{E})$ with binary-valued node labels $\boldsymbol{y}$
- Energy function $E : \{0,1\}^n \to \mathbb{R}$ with

$$E'(\boldsymbol{y}) \;:=\; \sum_{i\in\mathcal{V}} E'_i(y_i) \;+\; \sum_{(i,j)\in\mathcal{E}} E'_{ij}(y_i,y_j) \qquad (1)$$

Ising Model

- Edge cost E_{ij} is incurred only in states $\boldsymbol{y}$ where y_i and y_j disagree. Node energies are zero (*i.e.*, no *external field*).

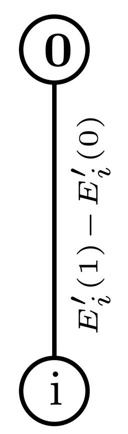
$$E(\boldsymbol{y}) \;:=\; \sum_{(i,j)\in\mathcal{E}} [y_i \neq y_j]\, E_{ij} \qquad (2)$$

Equivalence of (1) and (2)

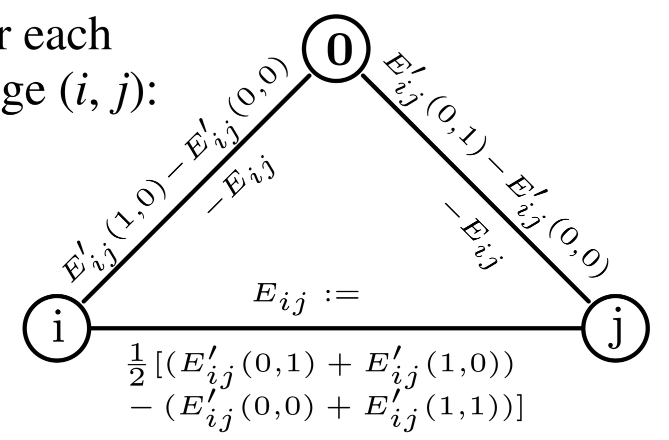
Theorem 1: (1) over n variables is equivalent to (2) over $n+1$ variables, with the additional variable held constant.

- Build Ising model with additional variable $y_0 := 0$ (bias node)
- Assign disagreement costs E_{ij} as shown:

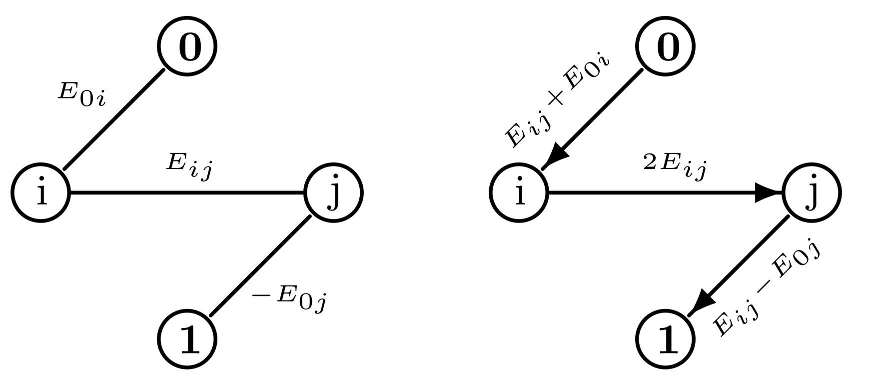
for each node i:



for each edge (i,j):



- For graph cuts, introduce another node $y_{n+1} := 1$, and reassign negative bias edges to it (left). Compare to directed graph construction of Kolmogorov/Zabih (2004) (right):



Planarity

- Graph cut methods depend on *submodularity*:

$$\forall (i,j) \in \mathcal{E}:\;\; E_{ij}(0,0) + E_{ij}(1,1) \le E_{i,j}(0,1) + E_{ij}(1,0)$$

- Instead we require the model graph to be *planar*:

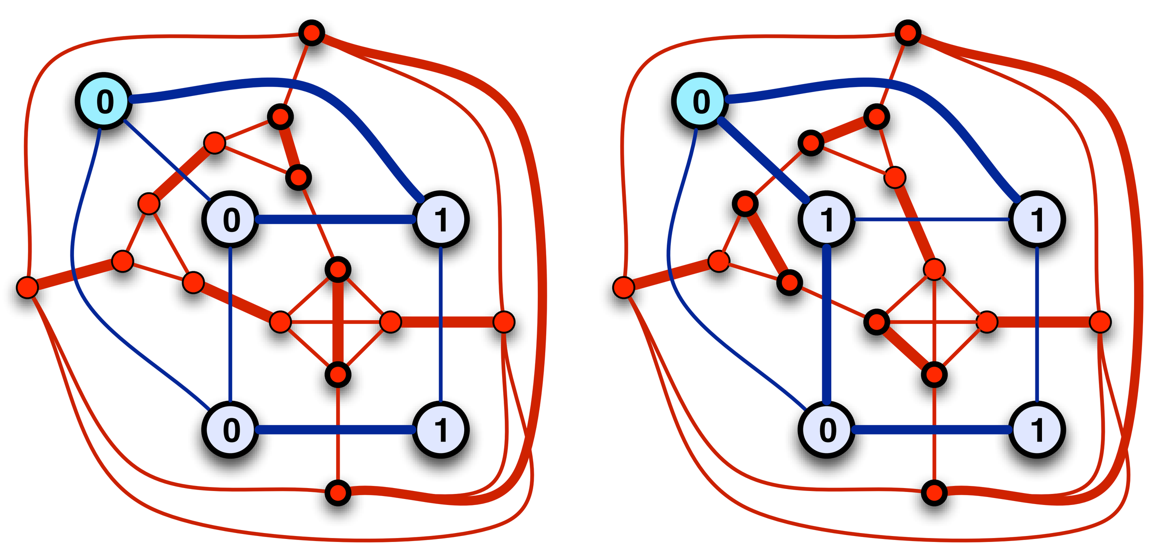
  Graph can be drawn in the plane $\mathbb{R}^2$ without crossing edges.

- Due to the planarity constraint, only nodes that lie on the same face (*e.g.*, perimeter of a grid) can be connected to bias node.

Computing Optimal States

Expanded Dual

- The *dual* $G^*(\mathcal{F},\mathcal{E})$ of an embedded graph $G(\mathcal{V},\mathcal{E})$ has a vertex for each face of G, with edges connecting vertices corresponding to adjacent faces in G.
- To obtain the *expanded dual* we replace each node in $G^*(\mathcal{F},\mathcal{E})$ with a q-clique, where q is the degree of the node:



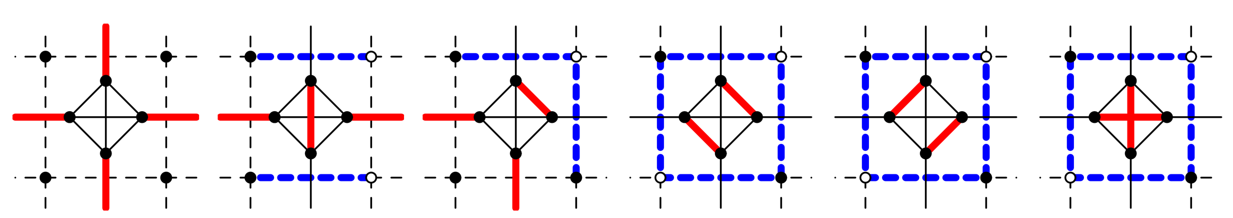
Model graph with 4 binary nodes & constant-valued bias node, and its expanded dual graph. Graph cut induced by given state and complementary perfect matching of the dual shown in bold.

Perfect Matchings and Graph Cuts

- The *cut* C of a binary graphical model $G(\mathcal{V},\mathcal{E})$ induced by state $\boldsymbol{y} \in \{0,1\}^n$ is the set $C(\boldsymbol{y}) := \{(i,j) \in \mathcal{E} : y_i \neq y_j\}$; its *weight* $|C(\boldsymbol{y})|$ is the sum of the weights of its edges.
- A *perfect matching* of a graph $G(\mathcal{V},\mathcal{E})$ is a subset $\mathcal{M} \subseteq \mathcal{E}$ of edges wherein exactly one edge is incident upon each vertex in $\mathcal{V}$; its *weight* $|\mathcal{M}|$ is the sum of the weights of its edges.

Theorem 2: For every cut C of an embedded graph $G(\mathcal{V},\mathcal{E})$ there is at least one (exactly one if G is triangulated) perfect matching $\mathcal{M}$ of its expanded dual *complementary* to $C : \mathcal{E}\backslash\mathcal{M} = C$.

Theorem 3: Every perfect matching $\mathcal{M}$ of the expanded dual of a *plane* graph $G(\mathcal{V},\mathcal{E})$ is complementary to a cut $C : \mathcal{E}\backslash\mathcal{M} = C$.



Possible cuts of a face of the model graph and complementary perfect matchings of its expanded dual.

Lowest-Energy (MAP) State

1. Construct the expanded dual of the model graph
2. Compute its *maximum-weight perfect matching* (via blossom-shrinking, takes $O(|\mathcal{E}|\,|\mathcal{V}|\log|\mathcal{V}|)$ time)
3. Its complement in the model graph is the *minimum-weight graph cut*, from which we can identify the MAP state

Partition Function and Marginals

- Markov Random Field (MRF) over (2) models distribution:

$$\mathbb{P}(\boldsymbol{y}) \;=\; \tfrac{1}{Z}\, e^{-E(\boldsymbol{y})}, \quad \text{where} \;\; Z := \sum_{\boldsymbol{y}} e^{-E(\boldsymbol{y})} \qquad (3)$$

- For *planar graphs*: $Z = \sqrt{|\boldsymbol{K}|}$. Our construction of $\boldsymbol{K}$:
  1. *plane triangulate* model graph
  2. *orient* the edges such that in-degree of every node is odd
  3. *construct* a Boolean matrix $\boldsymbol{H}$ from oriented graph
  4. *prefactor* the triangulation edges (from Step 1) out of $\boldsymbol{H}$
  5. build Kasteleyn matrix $\boldsymbol{K}$ from $\boldsymbol{H}$: add exponentiated disagreement costs along superdiagonal, skew-symmetrize
- The marginal probability of disagreement on the k^{th} edge:

$$\mathbb{P}(k \in C) \;=\; \boldsymbol{K}^{-1}_{2k-1,2k}\, \boldsymbol{K}_{2k-1,2k} \qquad (4)$$

CRF Parameter Estimation

- The disagreement costs in (2) are computed as $E_k := \boldsymbol{\theta}^{\top}\boldsymbol{x}_k$
- The *conditional* distribution $\mathbb{P}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$ modeled as a MRF (3)

Maximum-Likelihood

- L_2-regularized negative log likelihood and gradient w.r.t $\boldsymbol{\theta}$:

$$L_{\mathrm{ML}}(\boldsymbol{\theta}) \;:=\; \tfrac{1}{2}\lambda\|\boldsymbol{\theta}\|^2 + E(\boldsymbol{y}^*|\boldsymbol{x},\boldsymbol{\theta}) \;+\; \ln Z(\boldsymbol{\theta}|\boldsymbol{x})$$

$$\frac{\partial}{\partial\boldsymbol{\theta}} L_{\mathrm{ML}}(\boldsymbol{\theta}) \;=\; \lambda\boldsymbol{\theta} \;+\; \sum_{k\in\mathcal{E}} \Big([k \in C(\boldsymbol{y}^*)] \;-\; \mathbb{P}(k \in C(\boldsymbol{y}|\boldsymbol{x}))\Big)\, \boldsymbol{x}_k$$

- Marginally most likely (MPM) state: induced by the cut

$$\{k \in \mathcal{E} : \mathbb{P}(k \in C(\boldsymbol{y}|\boldsymbol{x})) > 0.5\}$$

Maximum-Margin

- *Worst margin violator* $\hat{\boldsymbol{y}} := \operatorname{argmin}_{\boldsymbol{y}} M(\boldsymbol{y}|\boldsymbol{y}^*_{;}\boldsymbol{x},\boldsymbol{\theta})$
- *Margin energy*:

$$M(\boldsymbol{y}|\boldsymbol{y}^*) \;:=\; E(\boldsymbol{y}) - \sum_{k\in\mathcal{E}} [[k \in C(\boldsymbol{y})] \;\neq\; [k \in C(\boldsymbol{y}^*)]]$$

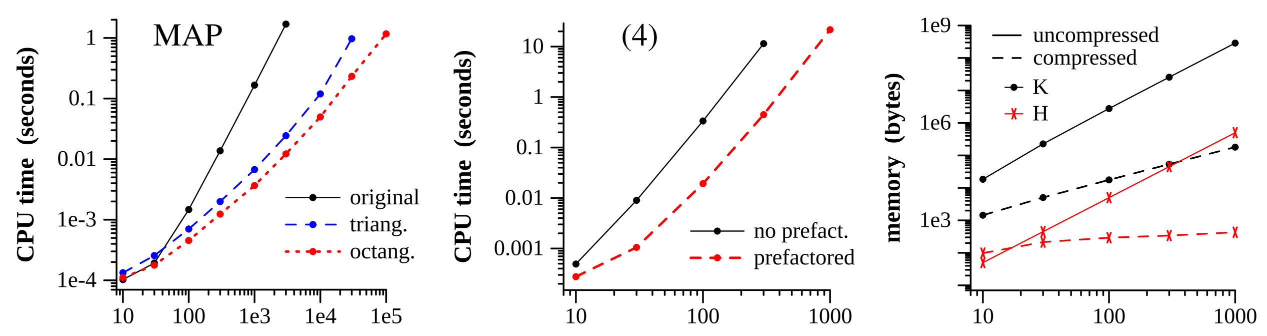
- L_2-regularized negative log likelihood and gradient w.r.t. $\boldsymbol{\theta}$:

$$L_{\mathrm{MM}}(\boldsymbol{\theta}) \;:=\; \tfrac{1}{2}\lambda\|\boldsymbol{\theta}\|^2 + E(\boldsymbol{y}^*|\boldsymbol{x},\boldsymbol{\theta}) - \min_{\boldsymbol{y}} M(\boldsymbol{y}|\boldsymbol{y}^*_{;}\boldsymbol{x},\boldsymbol{\theta})$$

$$\frac{\partial}{\partial\boldsymbol{\theta}} L_{\mathrm{MM}}(\boldsymbol{\theta}) \;=\; \lambda\boldsymbol{\theta} \;+\; \sum_{k\in\mathcal{E}} \Big([k \in C(\boldsymbol{y}^*)] \;-\; [k \in C(\hat{\boldsymbol{y}})]\Big)\, \boldsymbol{x}_k$$

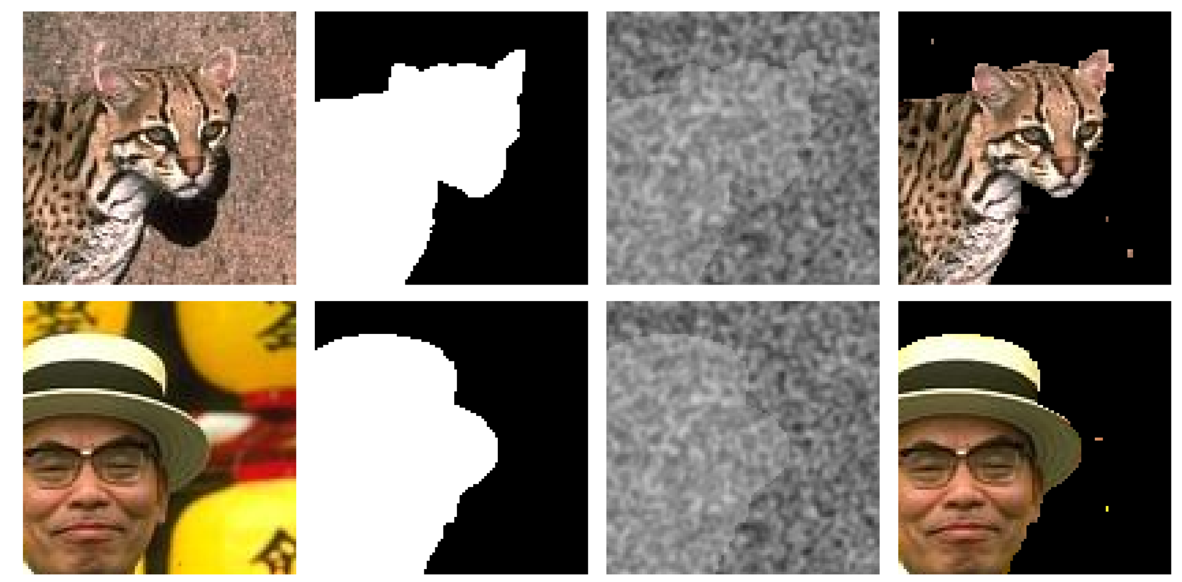
Results

Cost of inference



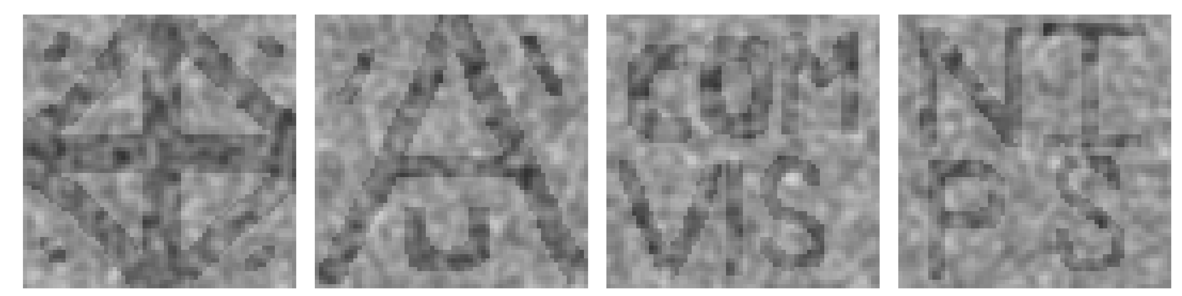
Cost of planar Ising inference methods on a ring graph, against ring size. $\boldsymbol{K}$ was compressed via row compression, $\boldsymbol{H}$ via bzip2.

Boundary Detection



Boundary detection, from left to right: original 100 × 100 image, original label, noisy input image, MAP segmentation.

Binary Denoising



(used for training)

Top row: noisy input image, bottom: MAP reconstruction.

Train Method	Patch Size	Train Time	Edge Error	Node Error
MM	64 × 64	490.4 s	1.15 %	2.10 %
	8 × 8	78.1 s	1.10 %	1.83 %
ML	8 × 8	5468.2 s	1.11 %	1.93 %

Method comparison; error differences statistically insignificant.

References

A. Globerson and T. Jaakkola, *Approximate inference using planar graph decomposition*, NIPS 19, pp. 473–480, 2007.

V. Kolmogorov and R. Zabih, *What energy functions can be minimized via graph cuts?*, IEEE Trans. PAMI 26(2):147–159, 2004.

N. N. Schraudolph and D. Kamenetsky. *Efficient exact inference in planar Ising models*, Technical Report 0810.4401, arXiv, 2008.