Tightness of LP Relaxations for Almost Balanced Models

Adrian Weller
University of Cambridge

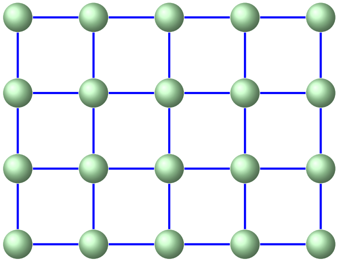
AISTATS
May 10, 2016

Joint work with Mark Rowland and David Sontag

For more information, see
http://mlg.eng.cam.ac.uk/adrian/

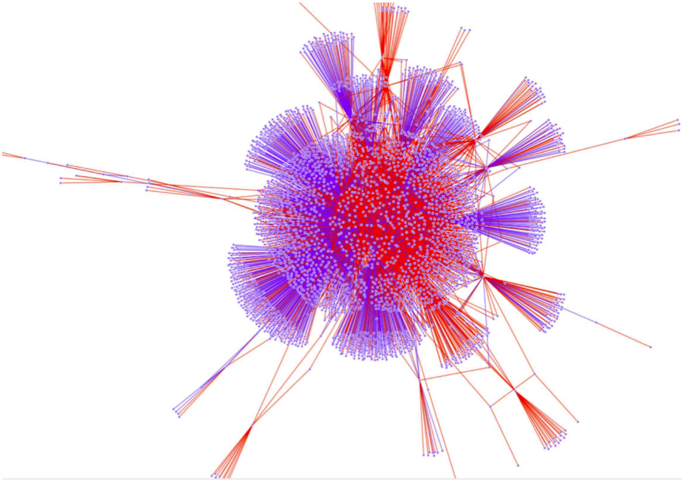
Motivation: *undirected graphical models*

- Powerful way to represent relationships across variables
- Many applications including: computer vision, social network analysis, deep belief networks, protein folding...
- In this talk, focus on binary pairwise (Ising) models



Example: Grid for computer vision (attractive)

Motivation: *undirected graphical models*



Example: Part of epinions social network

Figure courtesy of N. Ruozzi

Motivation: *undirected graphical models*

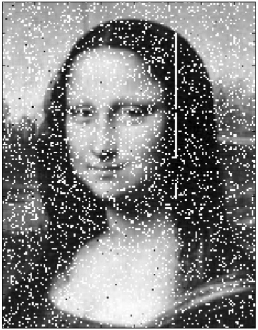
A fundamental problem is *maximum a posteriori (MAP) inference*

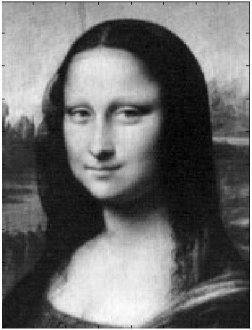
- Find a global configuration with highest probability

$$(x_1, \ldots, x_n)^* \in \arg\max p(x_1, x_2, \ldots, x_n)$$

- Example: image denoising

image from NASA



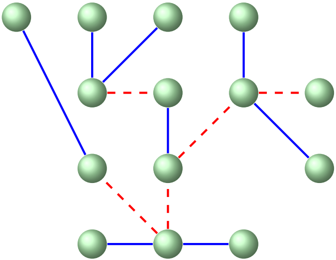


$\longrightarrow$ MAP inference

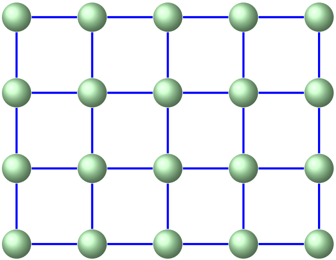
- Exponential search space, NP-hard in general

When is MAP inference (relatively) easy?

Tree



Attractive model



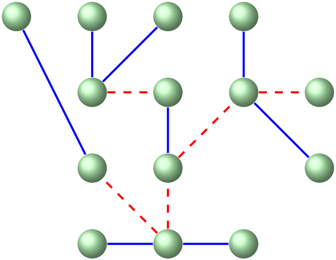
STRUCTURE

POTENTIALS

submodular costs

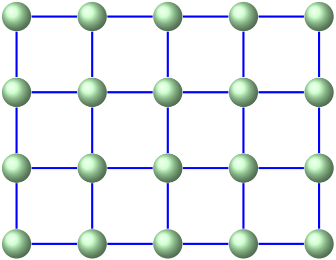
When is MAP inference (relatively) easy?

Tree



STRUCTURE

Attractive model



POTENTIALS

- Both can be solved exactly and efficiently with standard linear programming relaxation (LP+LOC): integer solution (tight)
- For models which are not attractive but are 'close to attractive', LP+LOC is often not tight - but using an LP relaxation with higher order clusters, empirically the result is tight (Sontag et al., 2008)

Example: Image foregound-background segmentation

(Domke, 2013)

- Learning potentials from data, most edges are attractive but a few are repulsive: the model is 'close to attractive'
- LP+LOC enforces pairwise consistency, often not tight
- The LP relaxation over the triplet polytope TRI usually is tight

Why?

Example: Image foregound-background segmentation

(Domke, 2013)

- Learning potentials from data, most edges are attractive but a few are repulsive: the model is 'close to attractive'
- LP+LOC enforces pairwise consistency, often not tight
- The LP relaxation over the triplet polytope TRI usually is tight

LP+TRI is tight for any almost attractive model

Example: Image foregound-background segmentation

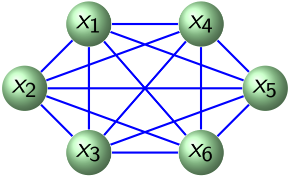
(Domke, 2013)

- Learning potentials from data, most edges are attractive but a few are repulsive: the model is 'close to attractive'
- LP+LOC enforces pairwise consistency, often not tight
- The LP relaxation over the triplet polytope TRI usually is tight

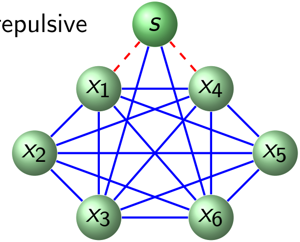
LP+TRI is tight for any almost balanced model

Almost attractive and almost balanced models

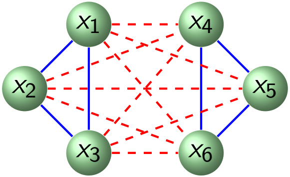
Blue edges are attractive, dashed red edges are repulsive



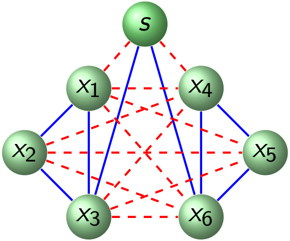
attractive



almost attractive



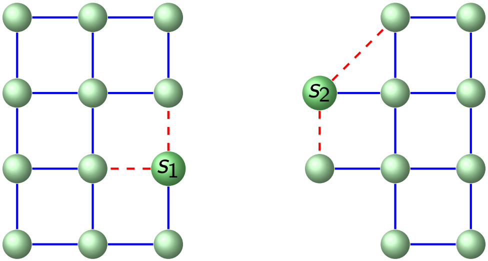
balanced
(attractive up to flipping)



almost balanced

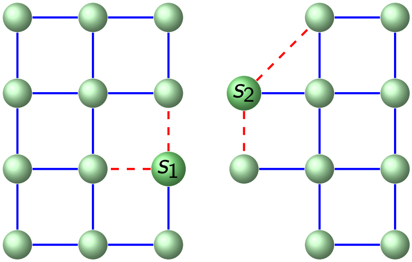
Main Results

- LP+TRI is tight for any almost balanced model
- We show a general result that submodels can be pasted together in certain ways while preserving LP tightness
- For LP+TRI
  - Can paste submodels on any one variable
  - Can paste on an edge provided it uses special variable s from each submodel



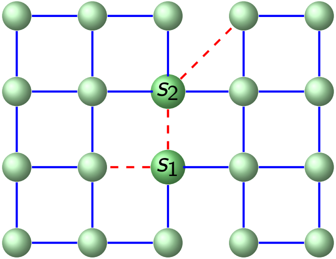
Main Results

- LP+TRI is tight for any almost balanced model
- We show a general result that submodels can be pasted together in certain ways while preserving LP tightness
- For LP+TRI
  - Can paste submodels on any one variable
  - Can paste on an edge provided it uses special variable s from each submodel



Main Results

- LP+TRI is tight for any almost balanced model
- We show a general result that submodels can be pasted together in certain ways while preserving LP tightness
- For LP+TRI
  - Can paste submodels on any one variable
  - Can paste on an edge provided it uses special variable s from each submodel



not almost balanced

Background: *Binary pairwise models, LP relaxations*

- Binary variables $X_1, \ldots, X_n \in \{0, 1\}$
- $p(x_1, \ldots, x_n) \propto \exp[\ \text{score}(x_1, \ldots, x_n)\ ]$ $\quad\leftarrow$ maximize
- $\text{score}(x_1, \ldots, x_n) = \sum_{i \in V} \theta_i x_i + \sum_{(i,j) \in E} W_{ij} x_i x_j$
- Singleton potentials: θ_i may take any value, often from data
- Edge potentials: $W_{ij} > 0$ attractive (supermodular potential, submodular cost); $W_{ij} < 0$ repulsive

- Combine singleton and edge potentials in a vector θ
- Write x for one 'complete configuration' of all variables,
  $\theta \cdot x$ for its score, contains singleton and edge terms

$$\theta = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_i \\ \ldots \\ \ldots \\ W_{ij} \\ \ldots \end{pmatrix} \qquad x = \begin{pmatrix} \mathbb{1}[X_1 = 1] \\ \ldots \\ \mathbb{1}[X_i = 1] \\ \ldots \\ \ldots \\ \mathbb{1}[X_i = 1, X_j = 1] \\ \ldots \end{pmatrix}$$

Background: *Binary pairwise models, LP relaxations*

- $\theta \cdot x$ is the score of a configuration x
- For MAP inference, now have a LP: $x^* \in \arg\max \theta \cdot x$

- Want to optimize over $\{0, 1\}$ coordinates of 'complete configuration space' corresponding to all 2^n possible settings
- The convex hull of these defines the marginal polytope $\mathbb{M}$, by construction has exactly these integral settings as its vertices
- Each point in $\mathbb{M}$ corresponds to a probability distribution over the 2^n configurations, giving a vector of marginals

- But optimizing over $\mathbb{M}$ is intractable: relax the space to pseudo-marginals q that enforce only local consistency, introduces fractional vertices

LOC and TRI polytopes

Recap

- Maximize $\theta \cdot q = \sum_{i \in V} \theta_i q_i + \sum_{(i,j) \in E} W_{ij} q_{ij}$ over singleton $\{q_i\}$ and edge $\{q_{ij}\}$ pseudo-marginals
- Edge potentials: if $W_{ij} > 0$ then the edge is attractive

LOC enforces **pairwise consistency**

- Ensures that every pair of variables has a valid distribution, all consistent with each other
- This requires $\max(0, q_i + q_j - 1) \leq q_{ij} \leq \min(q_i, q_j)$

LOC and TRI polytopes

Recap

- Maximize $\theta \cdot q = \sum_{i \in V} \theta_i q_i + \sum_{(i,j) \in E} \boldsymbol{W_{ij} q_{ij}}$ over singleton $\{q_i\}$ and edge $\{q_{ij}\}$ pseudo-marginals
- Edge potentials: if $\boldsymbol{W_{ij} > 0}$ then the edge is **attractive**

LOC enforces **pairwise consistency**

- Ensures that every pair of variables has a valid distribution, all consistent with each other
- This requires $\max(0, q_i + q_j - 1) \leq \boldsymbol{q_{ij}} \leq \boldsymbol{\min(q_i, q_j)}$

LOC and TRI polytopes

Recap

- Maximize $\theta \cdot q = \sum_{i \in V} \theta_i q_i + \sum_{(i,j) \in E} W_{ij} q_{ij}$ over singleton $\{q_i\}$ and edge $\{q_{ij}\}$ pseudo-marginals
- Edge potentials: if $W_{ij} > 0$ then the edge is attractive

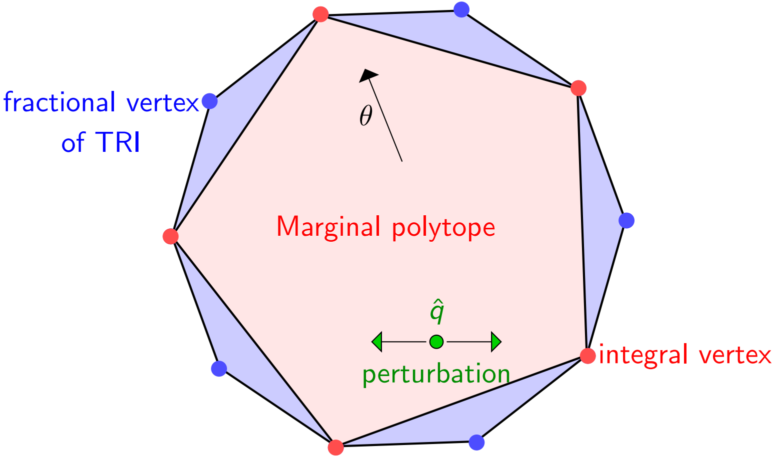
TRI enforces **triplet consistency**

- Ensures that every triplet of variables has a valid distribution, all consistent with each other
- This requires four additional inequalities for every triplet

Proof idea

Given an almost balanced model:

- if any non-integral optimum vertex $\hat{q}$ is proposed, we demonstrate an explicit small perturbation p s.t. $\hat{q} + p$ and $\hat{q} - p$ remain in TRI, while $\hat{q} = \frac{1}{2}(\hat{q} - p) + \frac{1}{2}(\hat{q} + p)$ and hence $\hat{q}$ cannot be a vertex



Key steps in the proof

- We may assume an almost attractive model: all edges are attractive except for some incident to variable s
- If s is held to a fixed marginal $q_s = y \in (0, 1)$, while all other marginals are optimized, some edge marginals 'behave as attractive edges' in LOC, i.e. $q_{ij} = \min(q_i, q_j)$
- We prove a structural result: any edge which is not 'behaving attractive' must be in a binding triplet constraint together with the special variable s

Key steps in the proof

- Given the structural result for fixed $q_s = y$, we construct an explicit perturbation up and down by p while remaining within TRI, unless all marginals take a simple form in $\{0, y, 1-y, 1\}$
- Hence at an optimum, all marginals must have this form
- We use this to show a stronger result:
  let $F^s(y) = \max_{q \in \text{TRI}: q_s = y} \theta \cdot q$ be the constrained optimum score in TRI holding fixed $q_s = y$, then $F^s(y)$ is linear
- Hence, the maximum is achieved at one end: $q_s = 0$ or $q_s = 1$
- Remaining model is attractive, hence global integer solution

Conclusion

- Previously known: LP+LOC is tight for attractive and balanced models
- Empirically LP relaxations using higher order cluster constraints are tight for models which are close to attractive
- We prove that LP+TRI is tight for almost attractive and almost balanced models
- We also provide a composition result
- This gives a **hybrid** condition on structure *and* potentials
- Connects to earlier work showing MAP inference is efficient for almost balanced models using perfect graphs (Weller, 2015)

Thank you

http://mlg.eng.cam.ac.uk/adrian/

References

- J. Domke. Learning graphical model parameters with approximate marginals inference. *TPAMI*, 2013.
- D. Sontag, T. Meltzer, A. Globerson, T. Jaakola and Y. Weiss. Tightening LP relaxations for MAP using message passing. In *UAI*, 2008.
- A. Weller. Revisiting the limits of MAP inference by MWSS on perfect graphs. In *AISTATS*, 2015.