Tightness of LP Relaxations for Almost Balanced Models

Mark Rowland

Joint work with Adrian Weller (Cambridge) and David Sontag (NYU)

9 September 2016

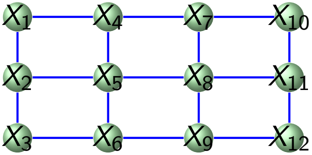
CP 2016

Originally presented at the International Conference on Artificial Intelligence and Statistics (AISTATS) 2016

For more information, see http://mr504.user.srcf.net/

Motivation: *Undirected graphical models*

Describe joint distribution of discrete random variables $X_1, \ldots, X_n$ according to graph $G = (V, E)$



$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \propto \exp\left(\sum_{\text{Cliques } C \text{ in } G} \psi_C(x_C) \right)$$

- Powerful way to represent relationships across variables
- Applications include computer vision, social network analysis, deep belief networks, protein folding...
- In this talk, focus on binary pairwise models

Motivation: *Undirected graphical models*

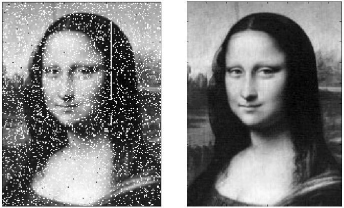
A fundamental problem is *maximum a posteriori (MAP) inference*

- Find a global configuration with highest probability

$$(x_1, \ldots, x_n)^* \in \arg\max \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

- Example: image denoising

image from NASA



$\longrightarrow$ MAP inference

- Exponential search space, NP-hard in general
- **One contribution: prove that this problem is tractable for a new class of models**

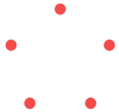
Motivation: *Undirected graphical models*

For a binary pairwise graphical model corresponding to $G = (V, E)$:

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \propto \exp\left(\sum_{i \in V} \theta_i \mathbb{1}_{x_i=1} + \sum_{ij \in E} W_{ij} \mathbb{1}_{x_i=1, x_j=1}\right)$$

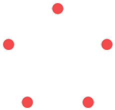
Combinatorial problem

$$\max_{x \in \{0,1\}^V} \left[\sum_{i \in V} \theta_i \mathbb{1}_{x_i=1} + \sum_{ij \in E} W_{ij} \mathbb{1}_{x_i=1, x_j=1}\right]$$



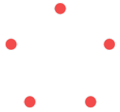
Combinatorial problem

$$\max_{x \in \{0,1\}^V} \left[\sum_{i \in V} \theta_i \mathbb{1}_{x_i=1} + \sum_{ij \in E} W_{ij} \mathbb{1}_{x_i=1, x_j=1} \right]$$



Combinatorial problem

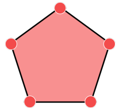
$$\max_{x\in\{0,1\}^V} \left[\sum_{i\in V} \theta_i \mathbb{1}_{x_i=1} + \sum_{ij\in E} W_{ij} \mathbb{1}_{x_i=1,x_j=1}\right]$$



Equivalent linear program

$$\max_{q\in\mathbb{M}} \left[\sum_{i\in V} \theta_i q_i + \sum_{ij\in E} W_{ij} q_{ij}\right]$$

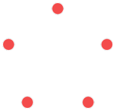
Marginal polytope $\mathbb{M}$: enforce global consistency on (pseudo)marginals $(q_i)_{i\in V}$ and $(q_{ij})_{ij\in E}$



Motivation: *Undirected graphical models*

Combinatorial problem

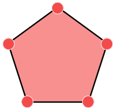
$$\max_{x \in \{0,1\}^V} \left[\sum_{i \in V} \theta_i \mathbb{1}_{x_i=1} + \sum_{ij \in E} W_{ij} \mathbb{1}_{x_i=1, x_j=1} \right]$$



Equivalent linear program

$$\max_{q \in \mathbb{M}} \left[\sum_{i \in V} \theta_i q_i + \sum_{ij \in E} W_{ij} q_{ij} \right]$$

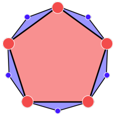
Marginal polytope $\mathbb{M}$: enforce global consistency on (pseudo)marginals $(q_i)_{i \in V}$ and $(q_{ij})_{ij \in E}$



Relaxed linear program

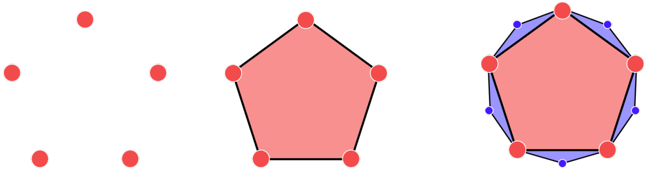
$$\max_{q \in \mathbb{L}_k} \left[\sum_{i \in V} \theta_i q_i + \sum_{ij \in E} W_{ij} q_{ij} \right]$$

Sherali-Adams polytope $\mathbb{L}_k$: enforce consistency over each cluster of k variables on (pseudo)marginals $(q_i)_{i \in V}$ and $(q_{ij})_{ij \in E}$



Motivation: *Undirected graphical models*

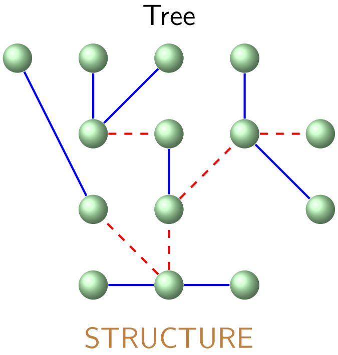
$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \propto \exp\left(\sum_{i \in V} \theta_i \mathbb{1}_{x_i=1} + \sum_{ij \in E} W_{ij} \mathbb{1}_{x_i=1, x_j=1} \right)$$

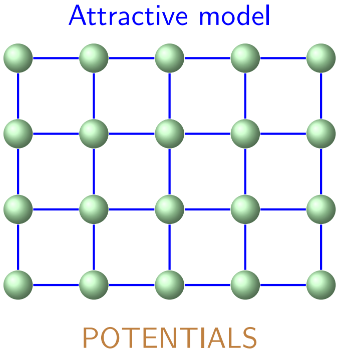


Combinatorial problem	Linear program over $\mathbb{M}$	Relaxed linear program over $\mathbb{L}_k$
Computationally hard	Computationally hard	**Computationally cheaper**
Exact	Exact	**Only sometimes exact**

- Marginal polytope $\mathbb{M}$: singleton and edge marginals $q = ((q_i)_{i \in V}, (q_{ij})_{ij \in E})$ which are *globally consistent*.
- Sherali-Adams polytope $\mathbb{L}_k$: singleton and edge pseudo-marginals that are *locally consistent* for each subset of k random variables. Common choices:
  - $\mathbb{L}_2$ - enforces consistency for all *pairs* of random variables
  - $\mathbb{L}_3$ - enforces consistency for *triplets* of random variables

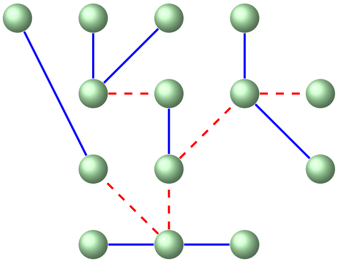
When is MAP inference (relatively) easy?





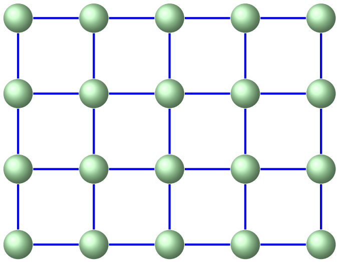
When is MAP inference (relatively) easy?

Tree



STRUCTURE

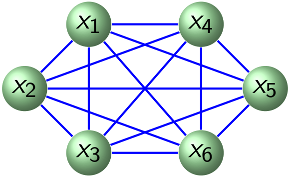
Attractive model



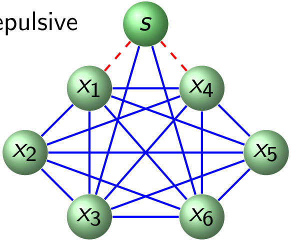
POTENTIALS

- Both can be solved exactly and efficiently over $\mathbb{L}_2$: integer solution (tight)
- For models which are not attractive but are 'close to attractive', $\mathbb{L}_2$ is often not tight - but using an LP relaxation with higher order clusters (e.g. $\mathbb{L}_3$), empirically the result is tight (Sontag et al., 2008)

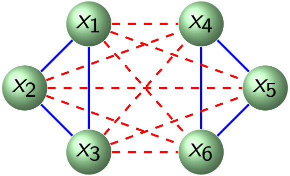
Blue edges are attractive, dashed red edges are repulsive



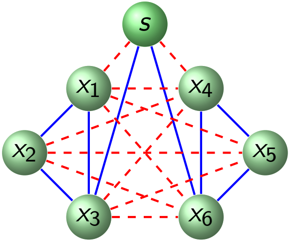
attractive



almost attractive



balanced
(attractive up to flipping)



almost balanced

Main Results

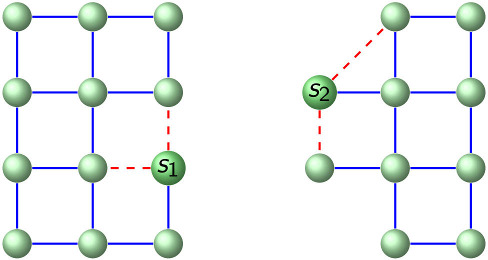
- $\mathbb{L}_3$ **is tight for any almost balanced model**

Main Results

- $\mathbb{L}_3$ **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**

Main Results

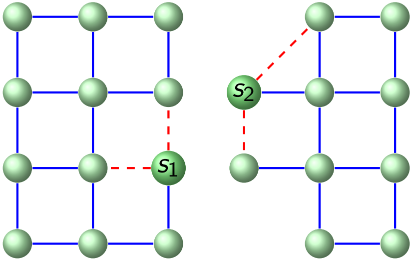
- $\mathbb{L}_3$ **is tight for any almost balanced model**
- **We show a general result that submodels can be pasted together in certain ways while preserving LP tightness**
- **For** $\mathbb{L}_3$**:**
  - **Can paste submodels on any one variable**
  - **Can paste on an edge provided it uses special variable** s **from each submodel**



not almost balanced, but $\mathbb{L}_3$ is tight

Main Results

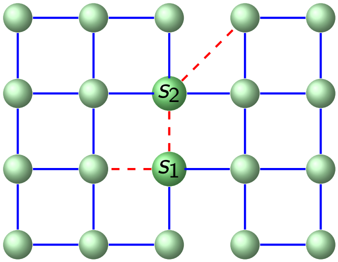
- $\mathbb{L}_3$ **is tight for any almost balanced model**
- **We show a general result that submodels can be pasted together in certain ways while preserving LP tightness**
- **For** $\mathbb{L}_3$**:**
  - **Can paste submodels on any one variable**
  - **Can paste on an edge provided it uses special variable** s **from each submodel**



not almost balanced, but $\mathbb{L}_3$ is tight

Main Results

- $\mathbb{L}_3$ **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** $\mathbb{L}_3$**:**
  - **Can paste submodels on any one variable**
  - **Can paste on an edge provided it uses special variable** s **from each submodel**

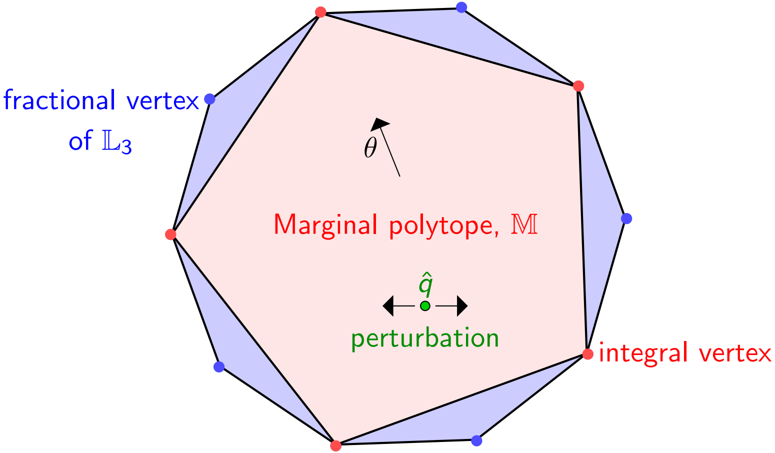


not almost balanced, but $\mathbb{L}_3$ is tight

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 $\mathbb{L}_3$, while $\hat{q} = \frac{1}{2}(\hat{q} - p) + \frac{1}{2}(\hat{q} + p)$ and hence $\hat{q}$ cannot be a vertex



Conclusion

- Previously known: $\mathbb{L}_2$ 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 $\mathbb{L}_3$ is tight for almost attractive and almost balanced models
- We also provide a composition result
- This gives a **hybrid** condition on structure *and* potentials

Thank you!

`http://mr504.user.srcf.net/`

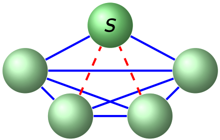
References

- 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.
- A. Weller, M. Rowland, D. Sontag. Tightness of LP Relaxations for Almost Balanced Models. In *AISTATS*, 2016.
- A. Weller. Characterizing Tightness of LP Relaxations by Forbidding Signed Minors. In *UAI*, 2016.

Extra slides for questions or further explanation

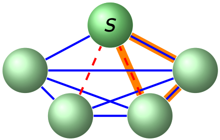
Key steps in the proof

We may assume an almost attractive model: all edges are attractive except for some incident to variable s



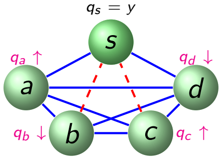
If s is held to a fixed marginal $q_s = y \in (0,1)$, while all other marginals are optimized, some edge marginals 'behave as attractive edges'

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



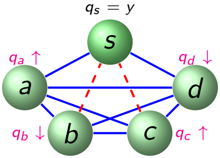
Key steps in the proof

Using 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 values in $\{0, y, 1-y, 1\}$.
Hence at an optimum, all marginals must have this form.

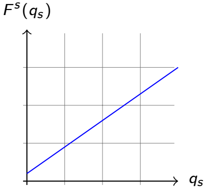


Key steps in the proof

Using 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 values in $\{0, y, 1-y, 1\}$.
Hence at an optimum, all marginals must have this form.

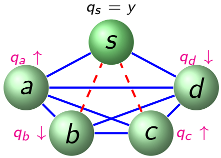


We use this to show a stronger result:
let $F^s(y) = \max_{q \in \mathbb{L}_3 : q_s = y} \theta \cdot q$ be the constrained optimum score in TRI holding fixed $q_s = y$, then $F^s(y)$ is linear.
Hence, the maximum is achieved at one end: $q_s = 0$ or $q_s = 1$.

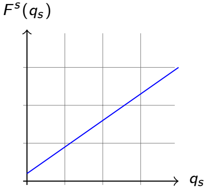


Key steps in the proof

Using 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 values 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 $F^s(y) = \max_{q \in \mathbb{L}_3 : q_s = y} \theta \cdot q$ 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.