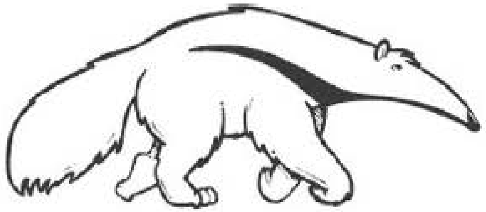
Bottom-up Approaches to Approximate Inference and Learning

Andrew Gelfand
Final Defense
April 10, 2014



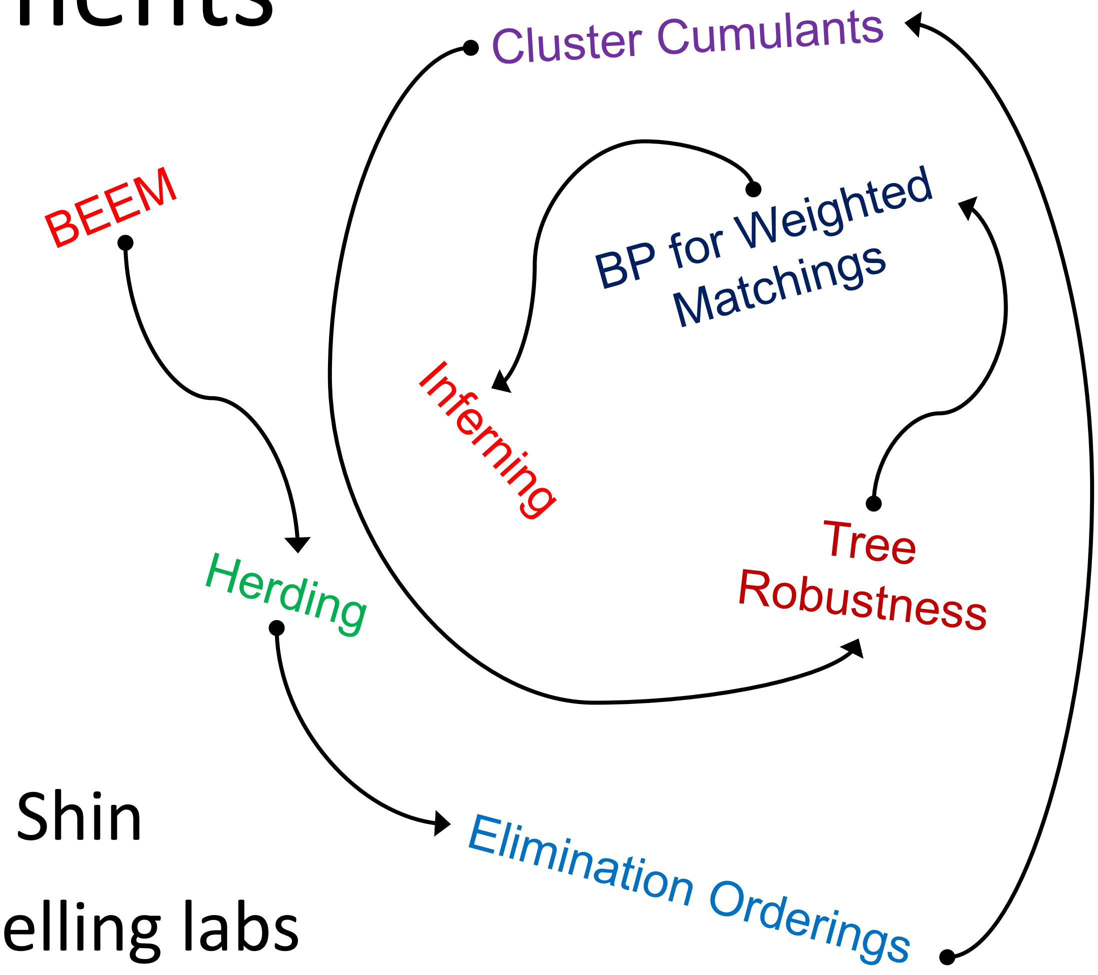
BREN:ICS INFORMATION AND COMPUTER SCIENCES

UNIVERSITY *of* CALIFORNIA IRVINE

Acknowledgements

- ☐ Rina Dechter
- ☐ Alex Ihler
- ☐ Max Welling
- ☐ Misha Chertkov

- ☐ Kalev Kask, Jinwoo Shin
- ☐ Dechter, Ihler & Welling labs



Cluster Cumulants

BEEM

BP for Weighted Matchings

Inferning

Tree Robustness

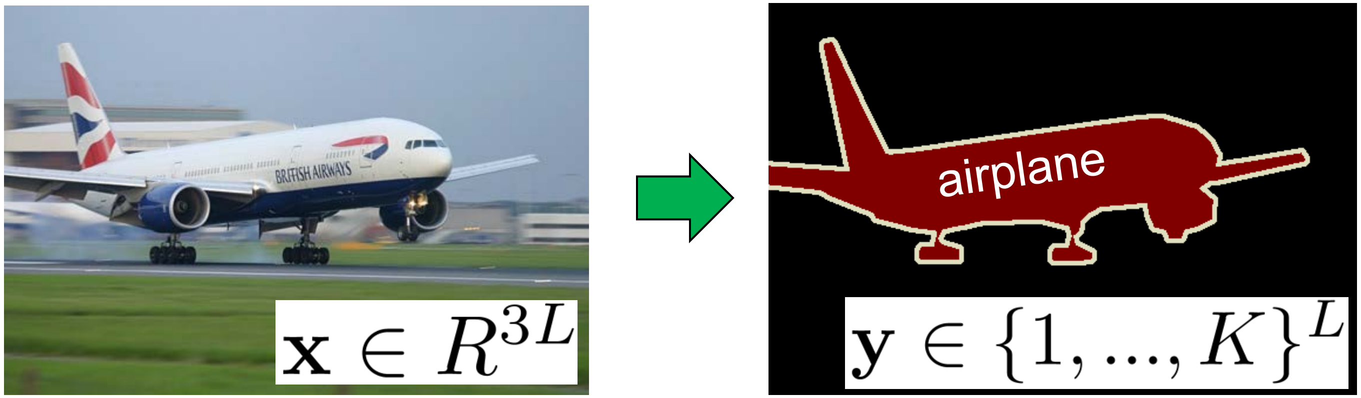
Herding

Elimination Orderings

Statistical Modeling

RGB image with *L* pixels

Labeling of *L* pixels



$\mathbf{x} \in R^{3L}$

$\mathbf{y} \in \{1, ..., K\}^L$

1. How do we represent $p(\mathbf{y}, \mathbf{x})$?
2. How do we learn $p(\mathbf{y}, \mathbf{x})$ from data?
3. How do we predict, *e.g.* compute $p\left(\mathbf{y} \middle| \right.$  $\left.\right)$?

Graphical Models

- Compact representation of large distributions
- Graph encodes probabilistic dependencies

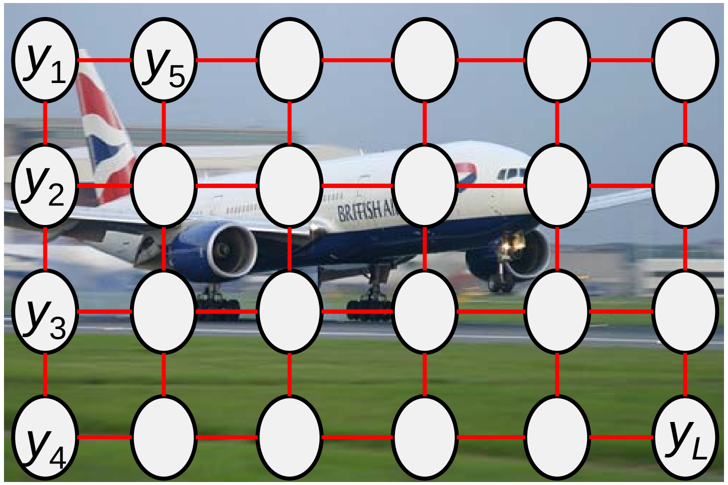
Ex: *Pairwise* model

$$p(\mathbf{y}|\mathbf{x}) \propto \prod_{i \in V} \psi_i(y_i, \mathbf{x}) \prod_{(i,j) \in E} \psi_{ij}(y_i, y_j, \mathbf{x})$$

$$y_i = \{\text{sky}, \text{plane}, \ldots, \text{car}\}$$

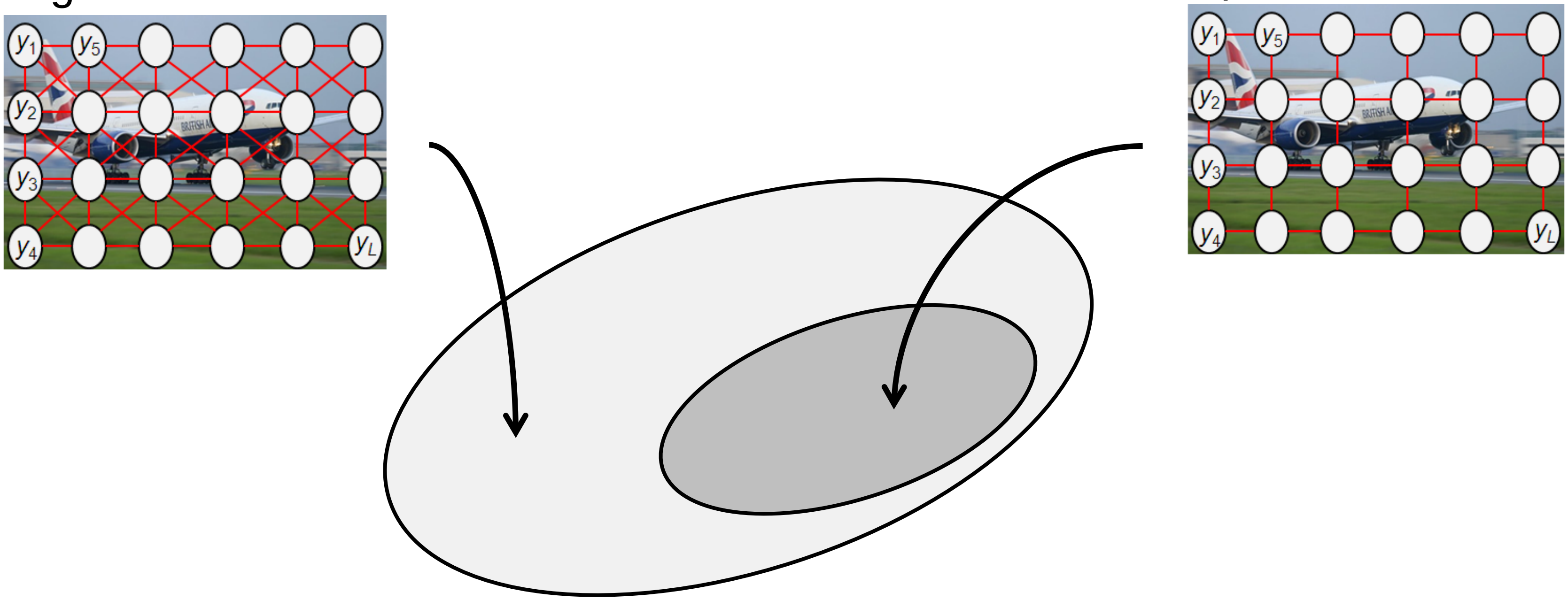
Only $|V|K + |E|K^2$ *parameters!*

$$G = (V, E)$$



Complexity of Graphical Models

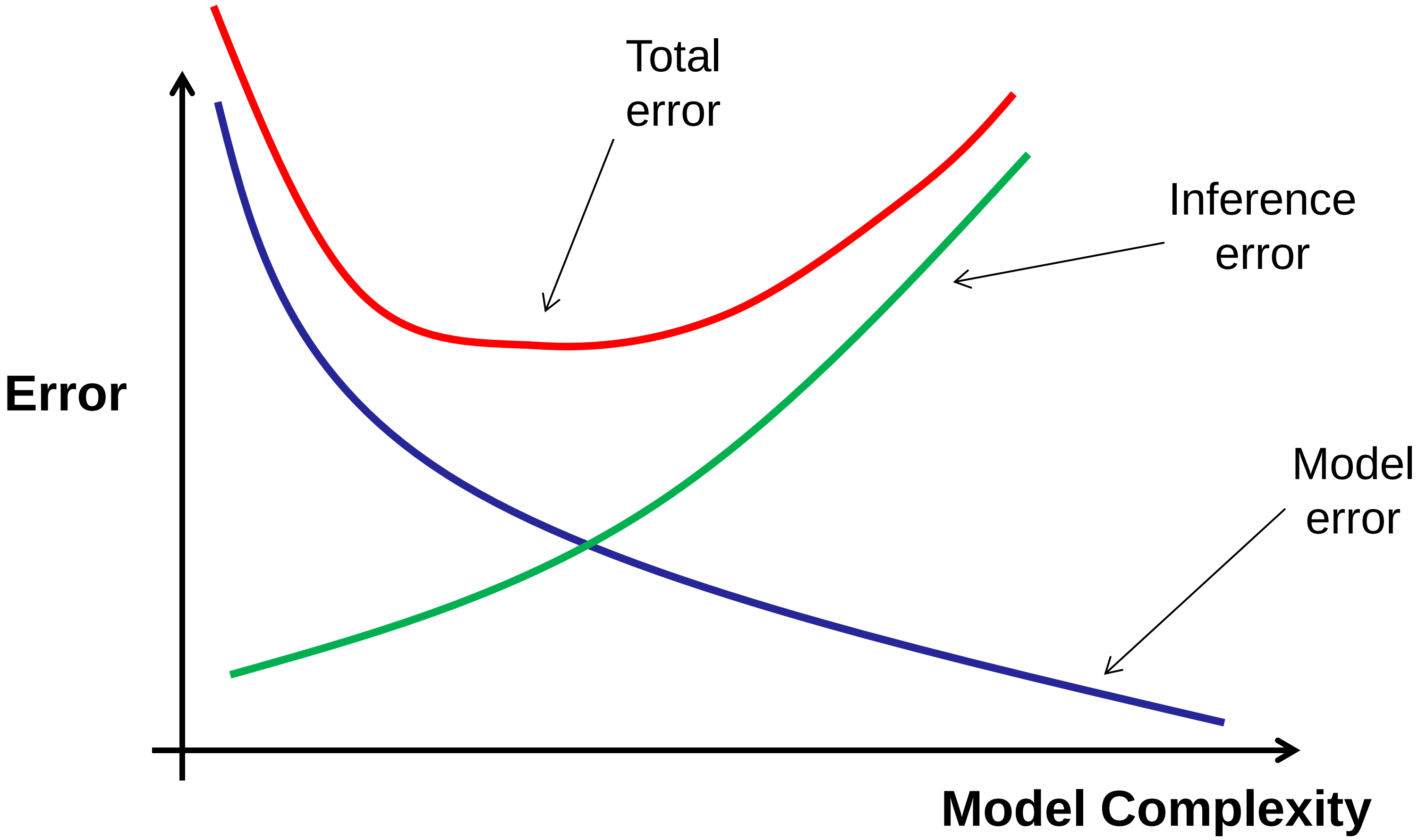
- A graphical model defines family of distributions



Higher-order model

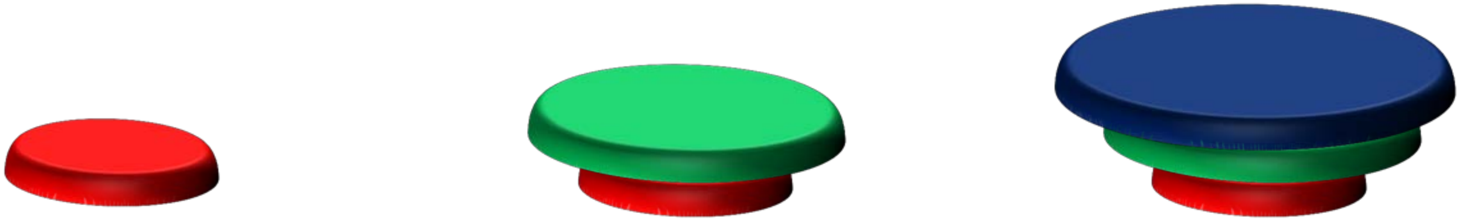
4-neighbor pairwise model

Are richer models more accurate?



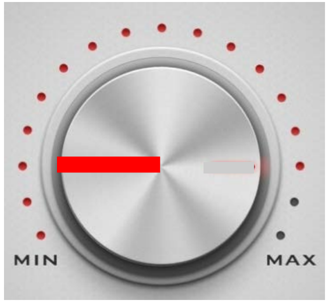
Thesis Statement

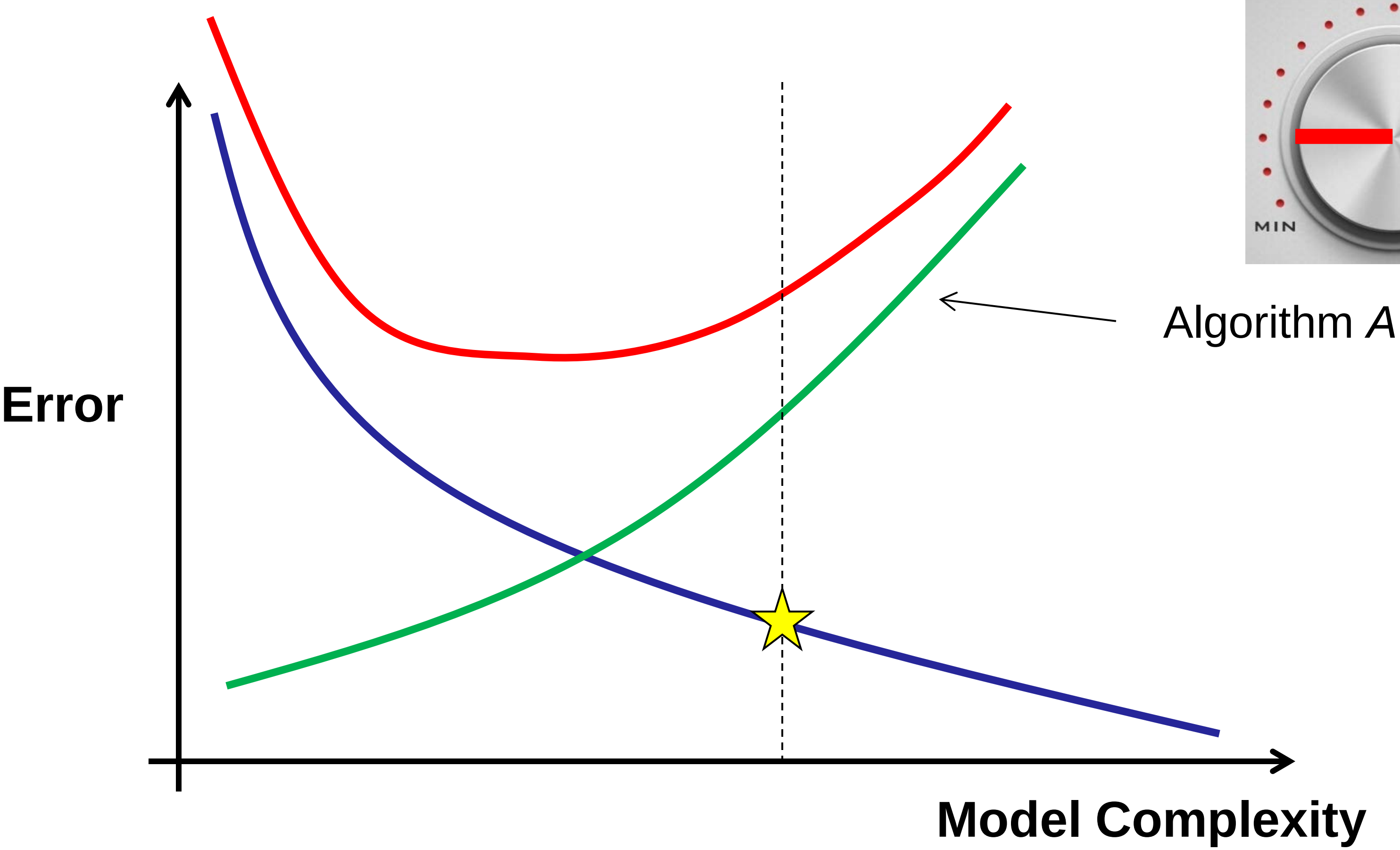
- Advocate a “bottom-up” approach to approximate inference & learning
  - Start with *simple*, cheap approximation
  - Improve through additional computation



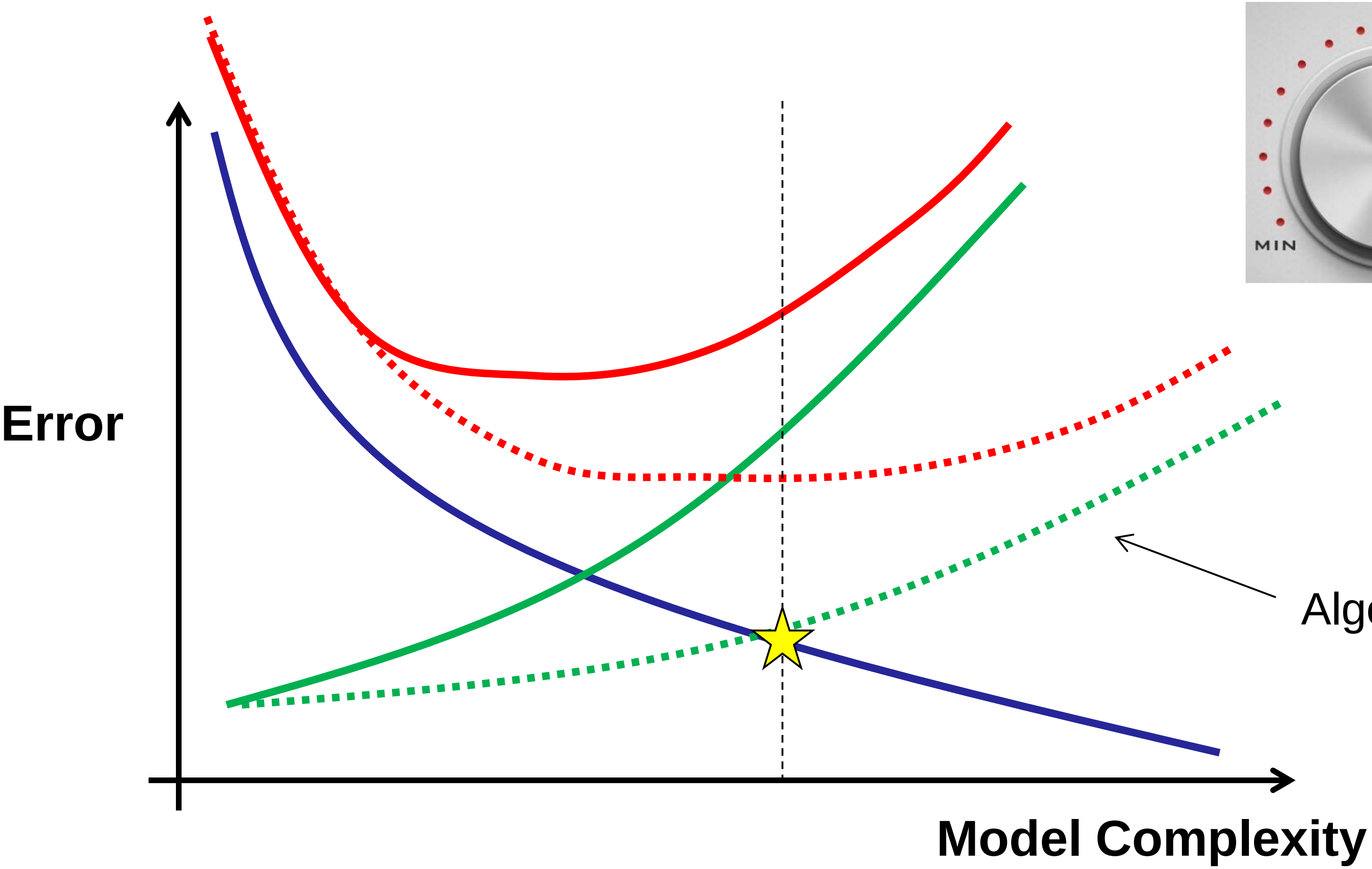
The “Bottom-up” approach

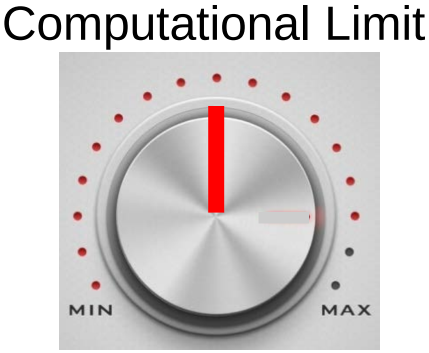
Computational Limit



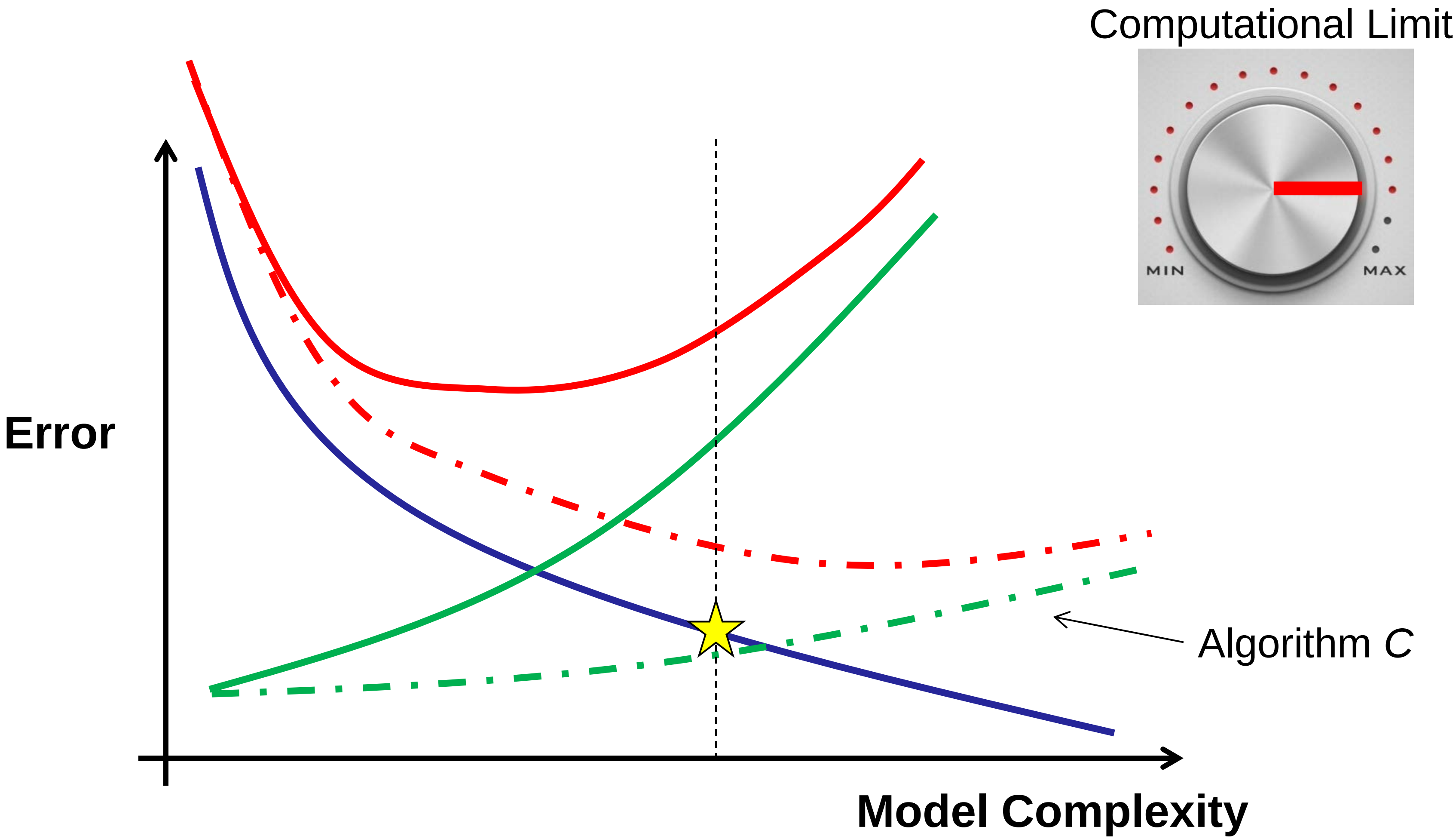


The “Bottom-up” approach





The “Bottom-up” approach



Overview of Thesis

1. Max Likelihood Learning
   - Computation-limited, approximate learning
2. Computing Marginal Probabilities
   - Region choice for Generalized Belief Propagation
3. Most Probable (MAP) Configuration
   - Cutting-plane algorithm for weighted matchings

Outline of this Talk

1. Max Likelihood Learning
    - Sources of error in likelihood-based learning
    - Computation-accuracy trade-offs in approximate learning
2. Computing Marginal Probabilities
    - Review of Belief Propagation (BP) & Generalized BP
    - Choosing Regions via Cycle Bases
3. Summary

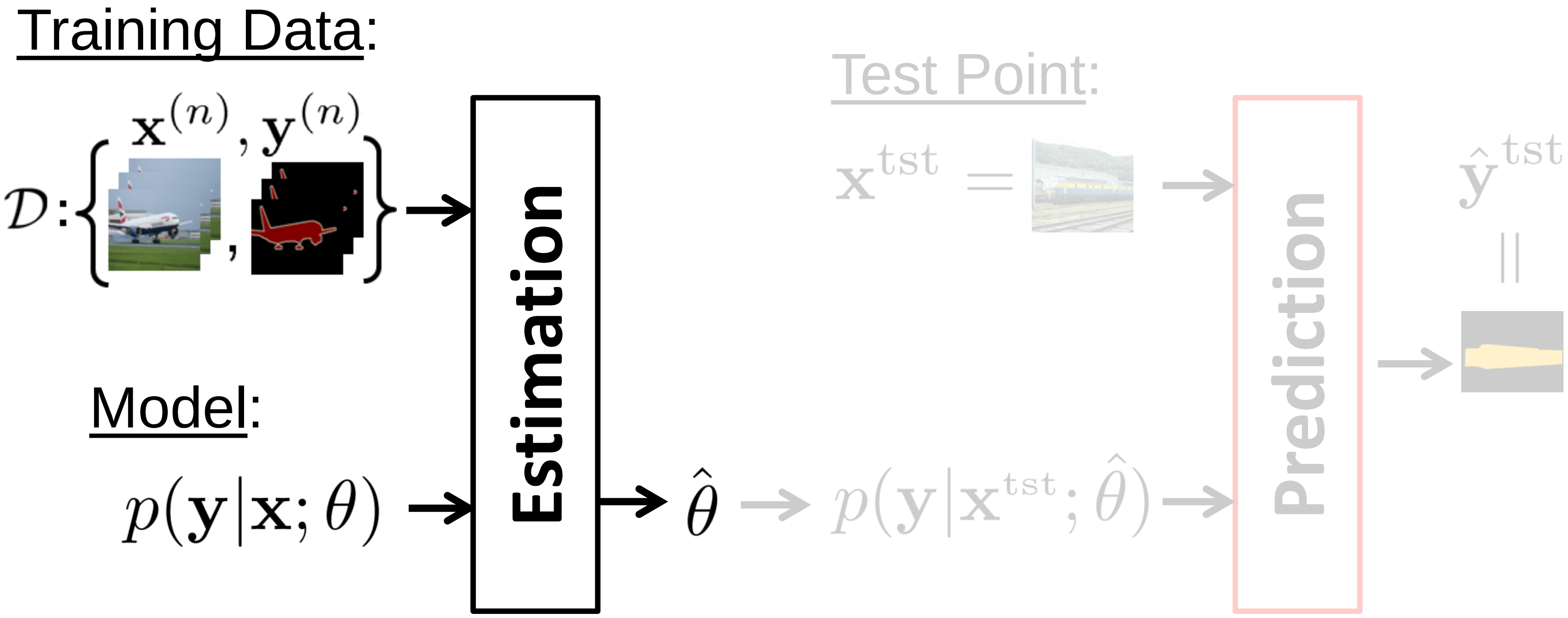
Outline of this Talk

1. Max Likelihood Learning
   - Sources of error in likelihood-based learning
   - Computation-accuracy trade-offs in approximate learning

2. Computing Marginal Probabilities
   - Review of Belief Propagation (BP) & Generalized BP
   - Choosing Regions via Cycle Bases

3. Summary

Parameter Estimation & Prediction



Training Data:

$\mathcal{D}: \left\{ \mathbf{x}^{(n)}, \mathbf{y}^{(n)} \right\}$

Model:

$p(\mathbf{y}|\mathbf{x};\theta)$

Estimation

$\hat{\theta} \rightarrow p(\mathbf{y}|\mathbf{x}^{\text{tst}};\hat{\theta})$

Test Point:

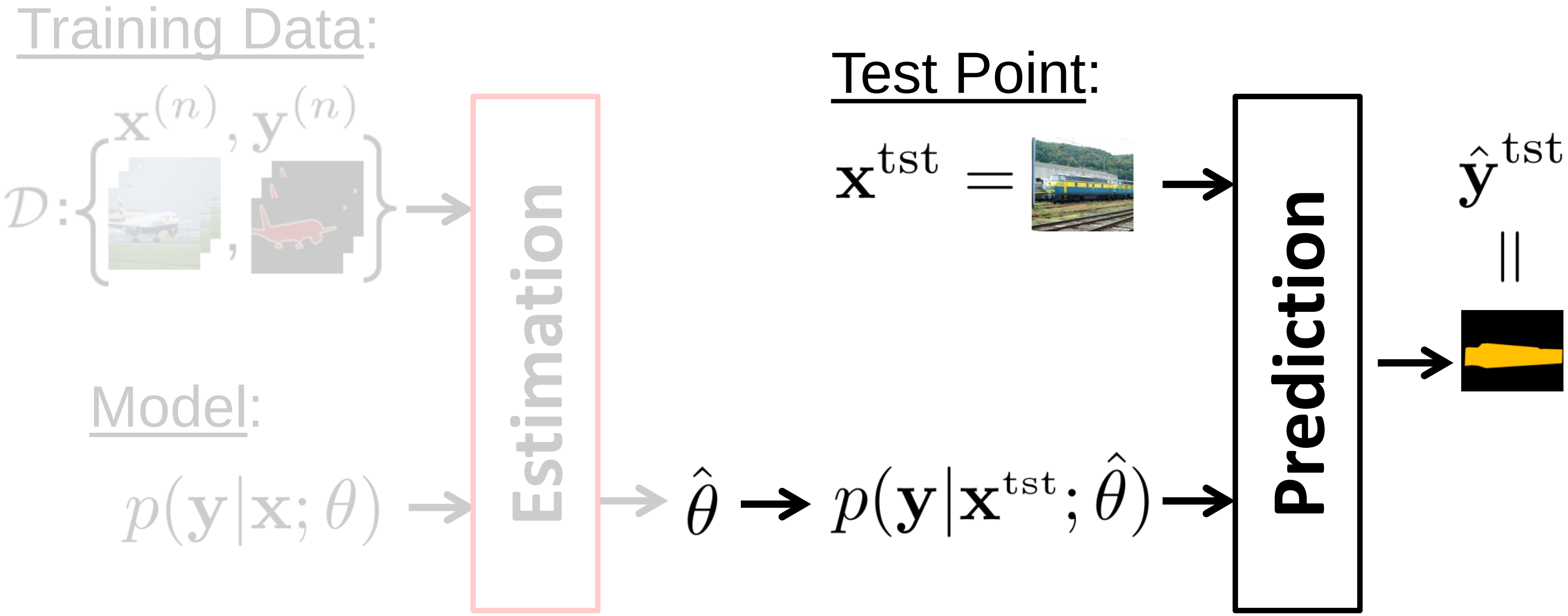
$\mathbf{x}^{\text{tst}} =$

Prediction

$\hat{\mathbf{y}}^{\text{tst}}$

$||$

Parameter Estimation & Prediction



Training Data:

$\mathcal{D}:\{\mathbf{x}^{(n)}, \mathbf{y}^{(n)}\}$

Model:

$p(\mathbf{y}|\mathbf{x};\theta)$

Estimation

$\hat{\theta} \rightarrow p(\mathbf{y}|\mathbf{x}^{\text{tst}};\hat{\theta})$

Test Point:

$\mathbf{x}^{\text{tst}} =$

Prediction

$\hat{\mathbf{y}}^{\text{tst}}$

$||$

Max Likelihood Estimation

- Given a model, $p(\boldsymbol{y};\boldsymbol{\theta}) = \exp\left(\sum_{c\in C}\theta_c(\boldsymbol{y}_c) - \log Z(\boldsymbol{\theta})\right)$,

  find

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^D} \ell_N(\boldsymbol{\theta})$$

  where,

$$\ell_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\log p(\boldsymbol{y}^{(n)};\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_N \cdot \boldsymbol{\theta} - \log Z(\boldsymbol{\theta})$$

vector of empirical marginals: $\bar{\mu}(\boldsymbol{y}_c) = \frac{1}{N}\sum_n I[\mathbf{Y}_c^{(n)} = \boldsymbol{y}_c]$

Surrogate Likelihood [Wainwright '06]

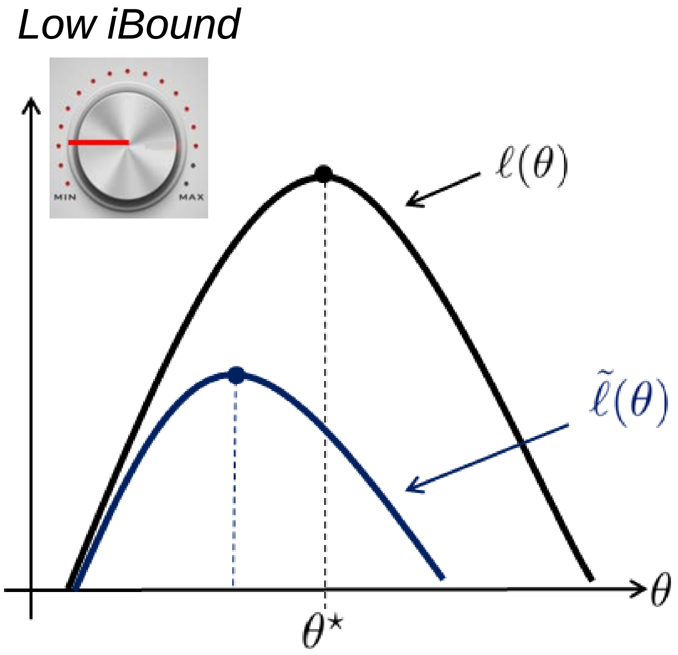
- Approximate $\log Z(\boldsymbol{\theta})$ with $\log \tilde{Z}(\boldsymbol{\theta})$,

$$\ell_N(\boldsymbol{\theta}) \approx \tilde{\ell}_N(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_N \cdot \boldsymbol{\theta} - \log \tilde{Z}(\boldsymbol{\theta})$$

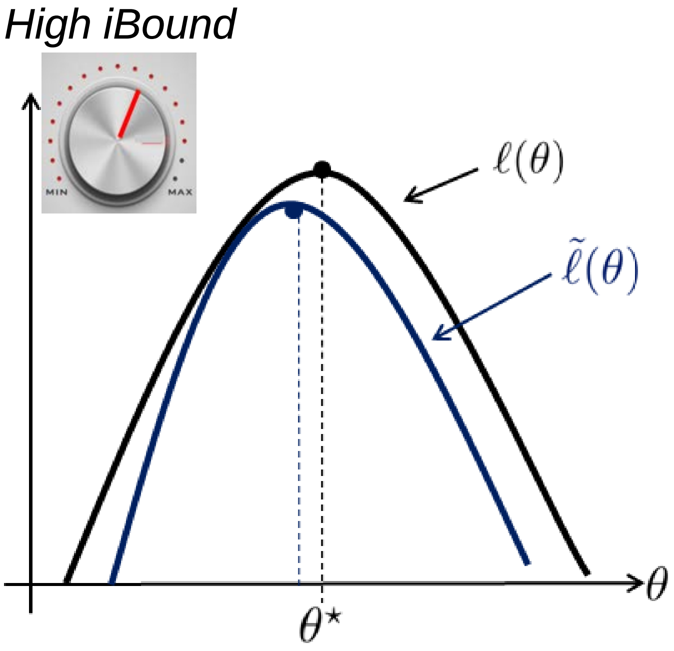
Surrogate Likelihood [Wainwright '06]

- Approximate $\log Z(\boldsymbol{\theta})$ with $\log \tilde{Z}(\boldsymbol{\theta})$,

$$\ell_N(\boldsymbol{\theta}) \approx \tilde{\ell}_N(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_N \cdot \boldsymbol{\theta} - \log \tilde{Z}(\boldsymbol{\theta})$$



, . . . ,

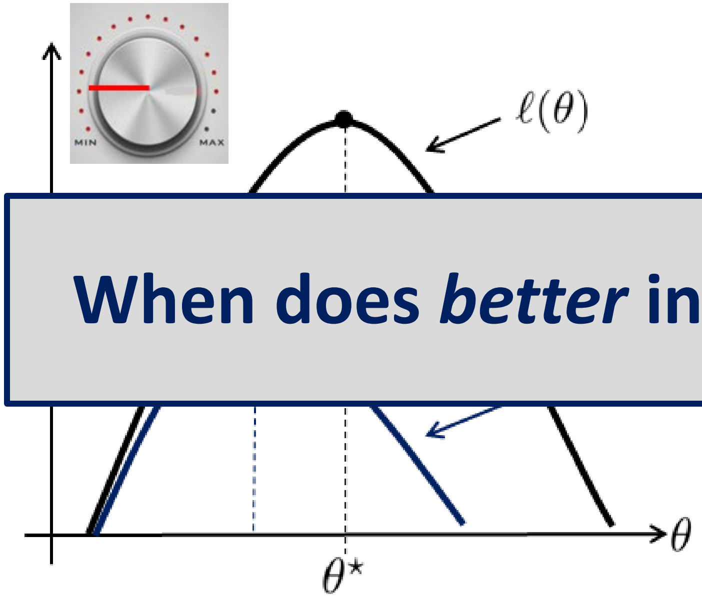


Surrogate Likelihood [Wainwright '06]

- Approximate $\log Z(\boldsymbol{\theta})$ with $\log \tilde{Z}(\boldsymbol{\theta})$,

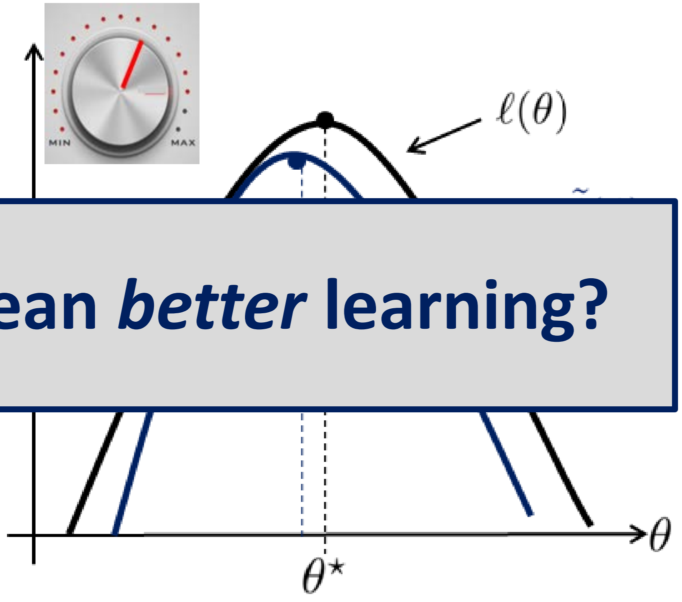
$$\ell_N(\boldsymbol{\theta}) \approx \tilde{\ell}_N(\boldsymbol{\theta}) = \bar{\boldsymbol{\mu}}_N \cdot \boldsymbol{\theta} - \log \tilde{Z}(\boldsymbol{\theta})$$

Low iBound



, . . . ,

High iBound



When does *better* inference mean *better* learning?

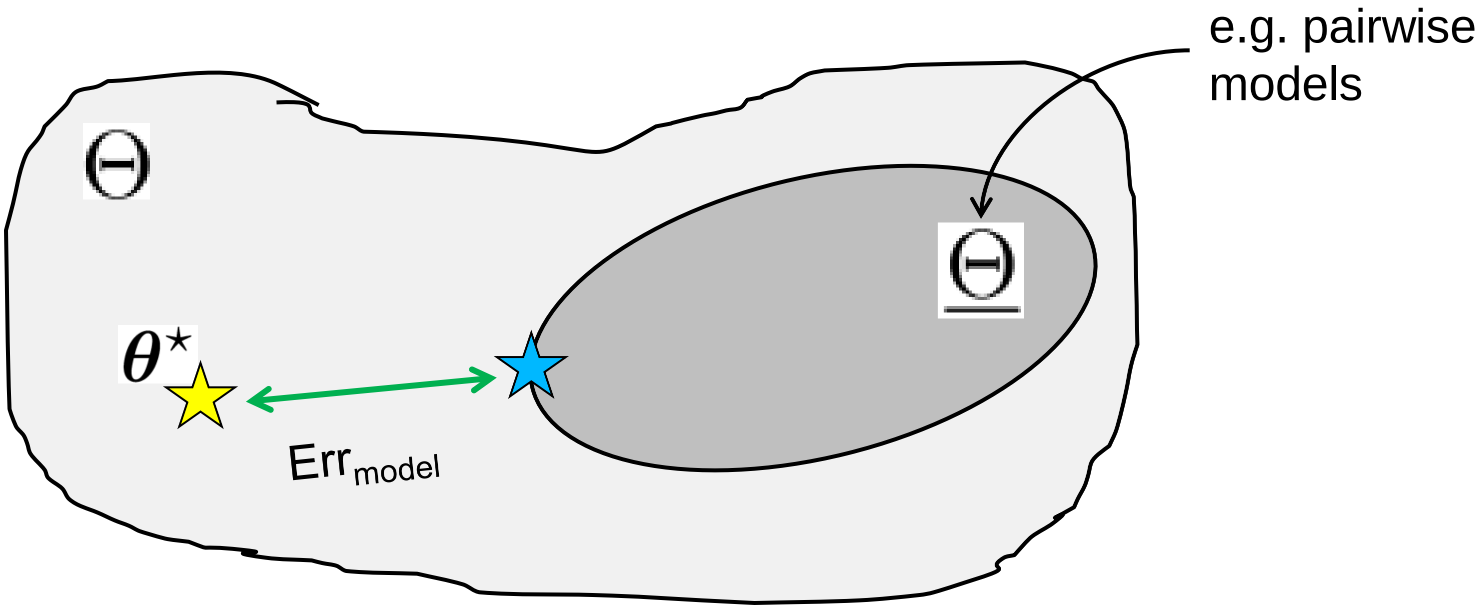
Errors in Approximate Learning

1. Model Error
   - Error in approximation to true (unknown) distribution
2. Estimation Error
   - Error due to use of finite sample
3. Optimization Error
   - *Gap* between true and surrogate likelihood functions

Errors in Approximate Learning

1. Model Error

- data sampled as $\boldsymbol{y}^{(n)} \overset{\text{iid}}{\sim} p(\boldsymbol{y}; \boldsymbol{\theta}^{\star})$, where $\boldsymbol{\theta}^{\star} \in \Theta$
- fit model with $\boldsymbol{\theta} \in \underline{\Theta}$

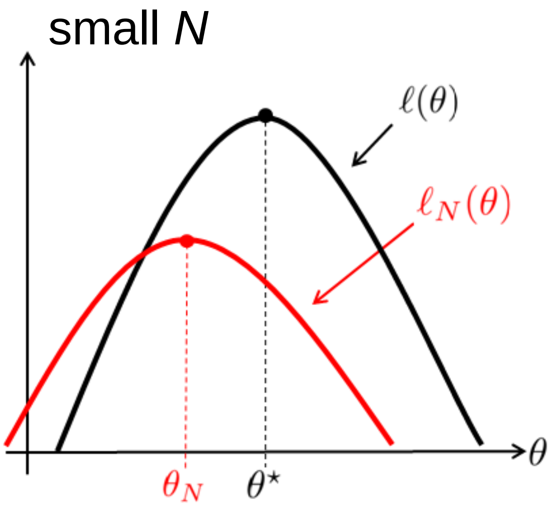


Errors in Approximate Learning

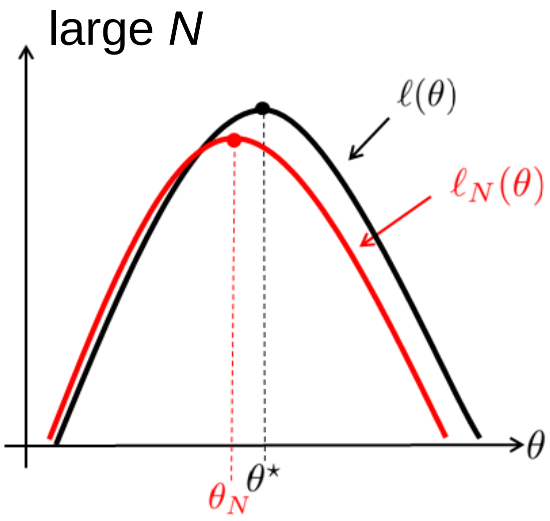
2. Estimation Error

- optimize empirical (not expected) risk

$$\ell_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\log p(\boldsymbol{y}^{(n)};\boldsymbol{\theta}) \overset{a.s.}{\to} \mathbb{E}_{\boldsymbol{\theta}^\star}\left[\log p(\boldsymbol{y};\boldsymbol{\theta})\right] = \sum_{\boldsymbol{y}} p(\boldsymbol{y};\boldsymbol{\theta}^\star)\log p(\boldsymbol{y};\boldsymbol{\theta})$$



, . . . ,

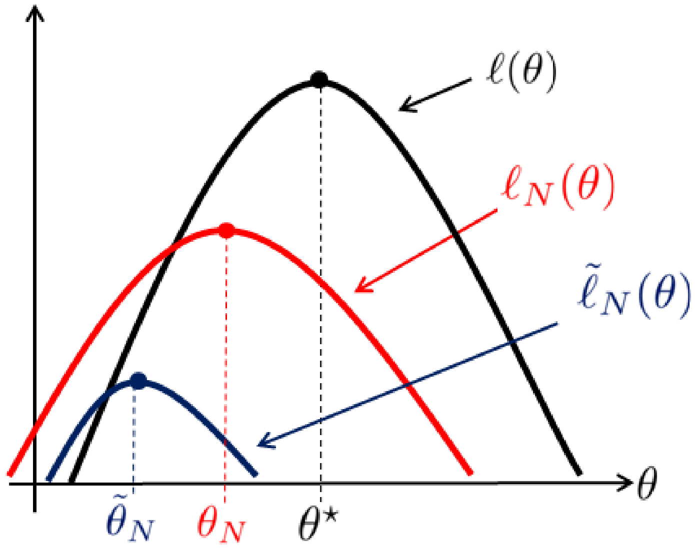


Errors in Approximate Learning

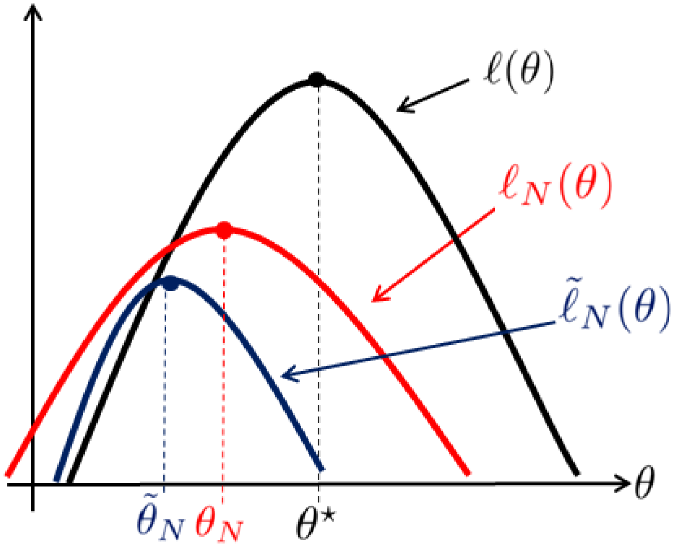
3. Optimization Error

- optimize a surrogate likelihood $\tilde{\ell}_N(\boldsymbol{\theta}) \approx \ell_N(\boldsymbol{\theta})$

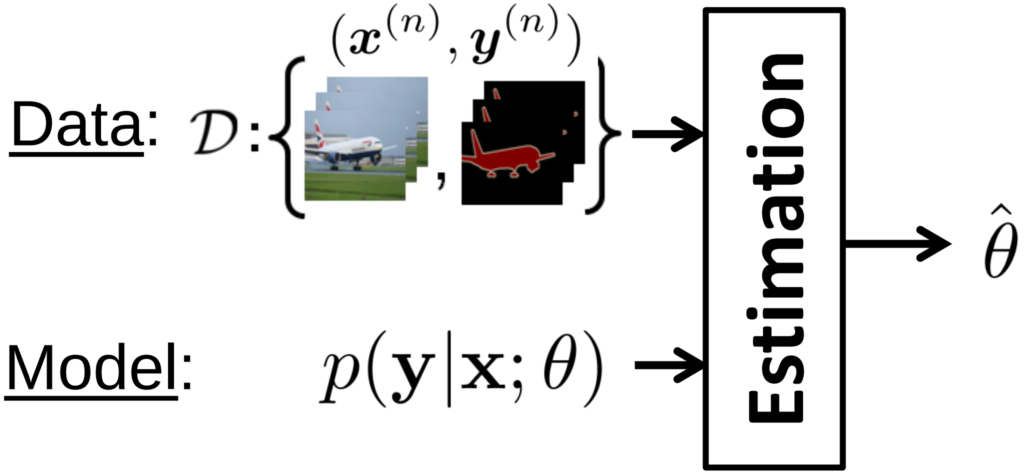
surrogate *A* (poor)



surrogate *B* (better)



Empirical Study (I) – Estimation



Data: $\mathcal{D}: \left\{ (\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) \right\} \rightarrow$

Model: $p(\mathbf{y}|\mathbf{x};\theta) \rightarrow$

$\rightarrow \hat{\theta}$

$$(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) \overset{\text{iid}}{\sim} p(\boldsymbol{y};\boldsymbol{\theta}^{\star})$$

$$\text{MSE}: \ ||\hat{\theta} - \theta^{\star}||^2$$

- Study estimation error (MSE) as we vary:
  - Optimization Error: inference procedures
  - Estimation Error: data set size
  - Model Error: mismatch between Θ and $\underline{\Theta}$

Experimental Setup

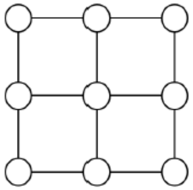
- Create *L* statistically identifiable models
  - Sample $\theta_i(y_i) \sim \mathcal{N}(0, \sigma_i^2)$ and $\theta_{ij}(y_i, y_j) \sim \mathcal{N}(0, \sigma_{ij}^2)$ for

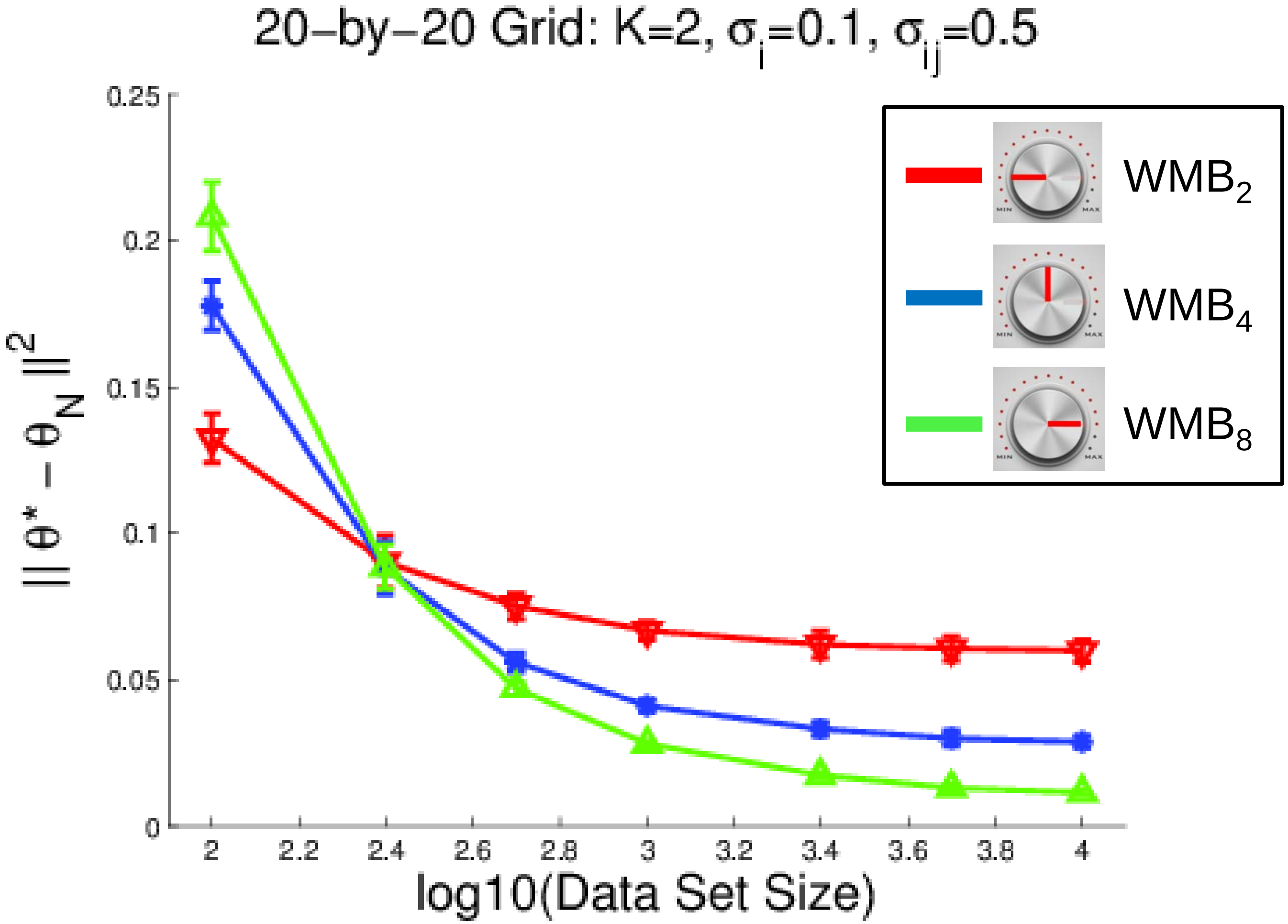
$$p(\boldsymbol{y}; \boldsymbol{\theta}) = \exp\left(\sum_{i \in V} \theta_i(y_i) + \sum_{(i,j) \in E} \theta_{ij}(y_i, y_j) - \log Z(\boldsymbol{\theta}) \right)$$

- Sample data sets of size *N* = {100,…,10000}
- Find $\tilde{\boldsymbol{\theta}}_N = \arg\max_{\boldsymbol{\theta}} \tilde{\ell}_N(\boldsymbol{\theta})$
  - Using different approximate inference algorithms

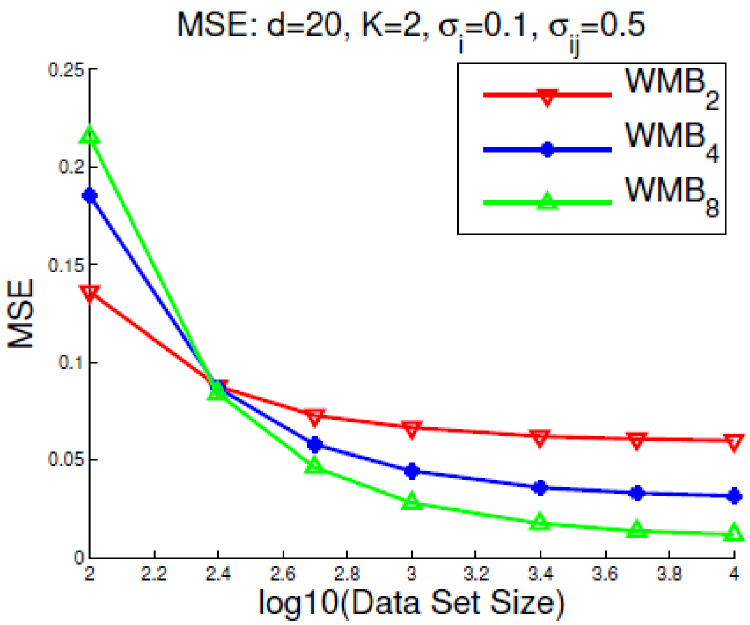
Grid Experiments

$\Theta = \underline{\Theta} =$

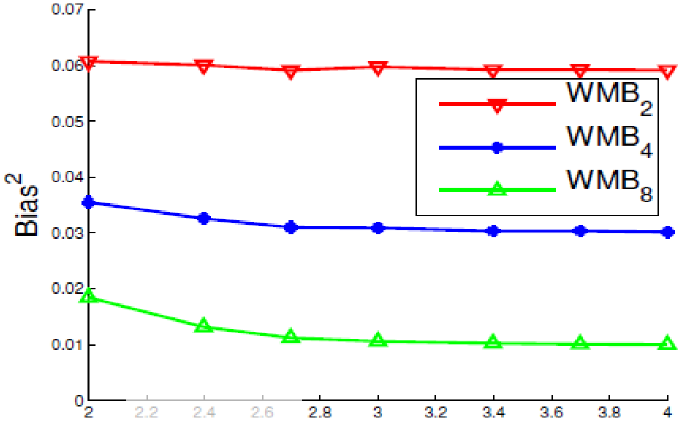




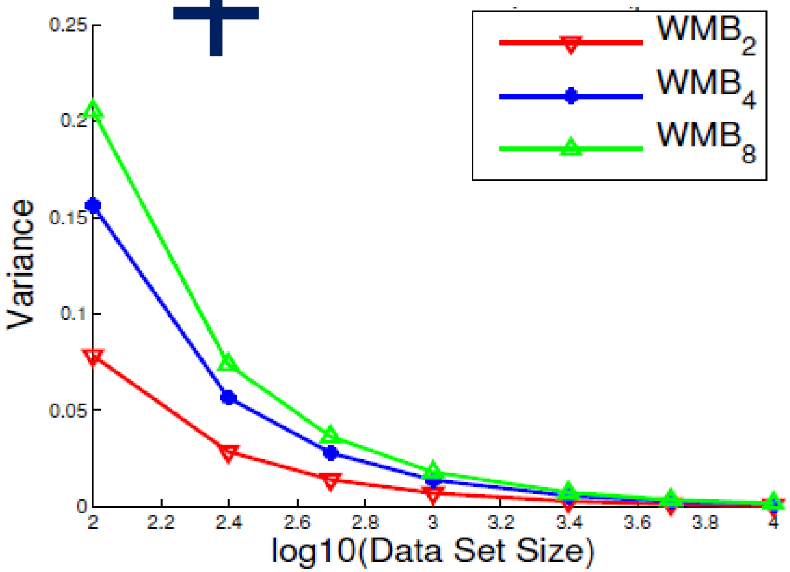
Bias–Variance Trade-off



=



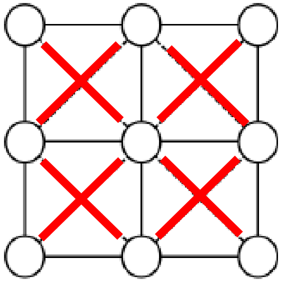
+



Model Error Experiments

Generate

$\Theta =$

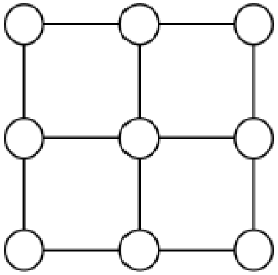


$$\boldsymbol{y}^{(n)} \overset{\text{iid}}{\sim} p(\boldsymbol{y};\boldsymbol{\theta}^\star) \quad \text{where } \boldsymbol{\theta}^\star \in \Theta$$

$$\theta_{ik}(y_i, y_k) \sim \mathcal{N}(0, \sigma_{ik}^2)$$

Estimate

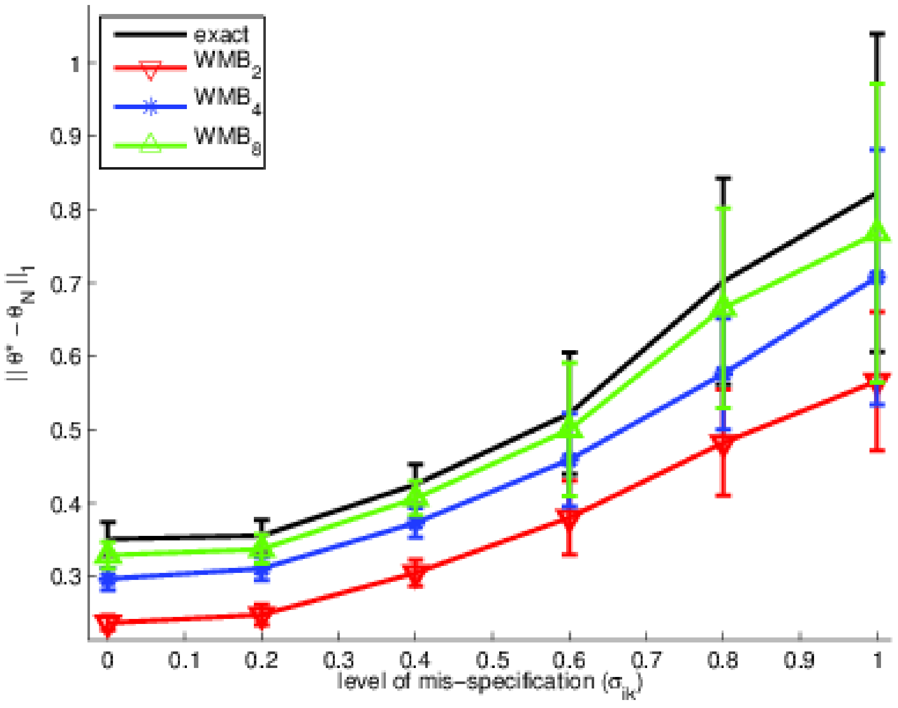
$\underline{\Theta} =$



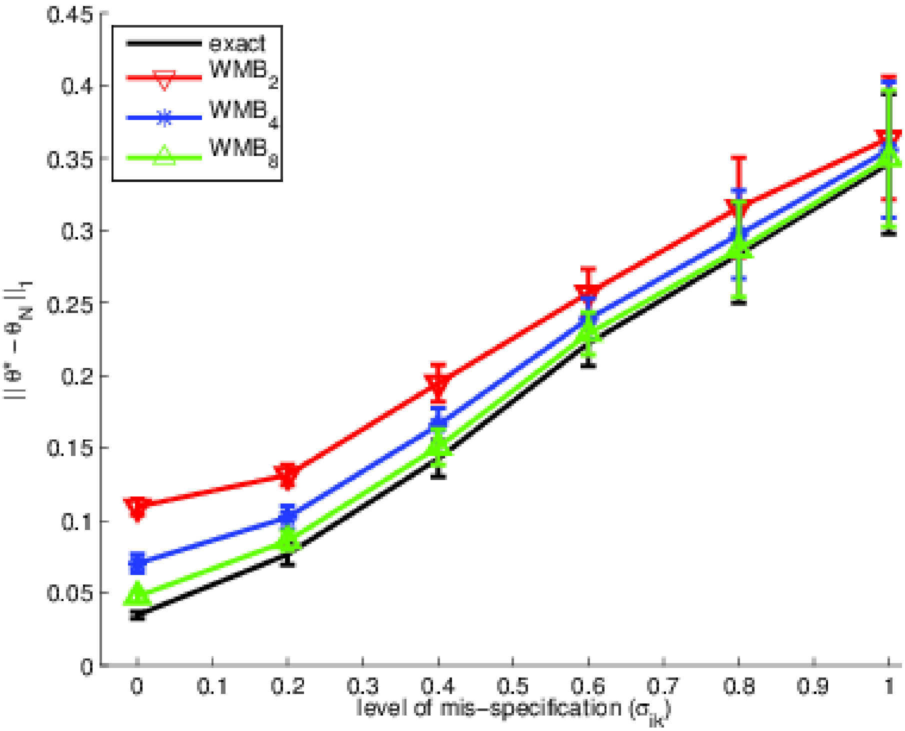
$$p(\boldsymbol{y};\boldsymbol{\theta}) \quad \text{where } \boldsymbol{\theta} \in \underline{\Theta}$$

Model Error Experiments

N=100

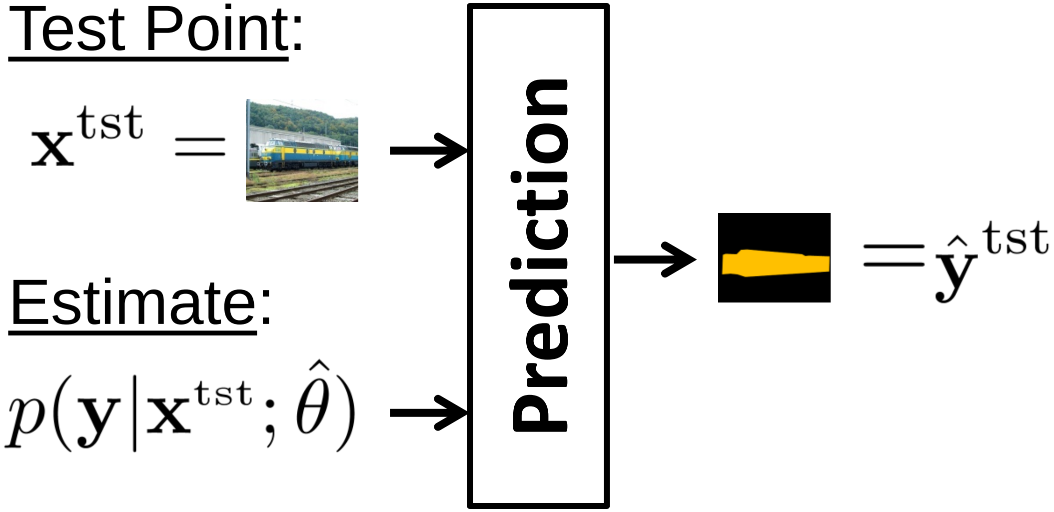


N=10000



→ Increasing amount of mis-specification

Empirical Study (II) – Prediction



Test Point:

$\mathbf{x}^{\mathrm{tst}} =$

Estimate:

$p(\mathbf{y} | \mathbf{x}^{\mathrm{tst}}; \hat{\theta})$

Prediction

$= \hat{\mathbf{y}}^{\mathrm{tst}}$

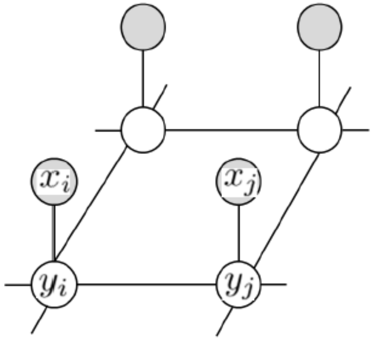
Typically don't know $\theta^{\star}$, compute

$$\underbrace{\mathrm{Error}_{\mathrm{pred}}(\hat{\mathbf{y}}^{\mathrm{tst}}, \mathbf{y}^{\mathrm{tst}})}_{\text{Hamming Loss: } \frac{1}{|V|} \sum_{i \in V} I\left[\hat{y}_i^{\mathrm{tst}} \neq y_i^{\mathrm{tst}}\right]}$$

MRF De-Noising Experiments

- USPS digits data (1100, 16x16 pixel grayscale images)
  - Convert to binary and flip pixels with probability *p*
- Assume a 4-neighbor model:

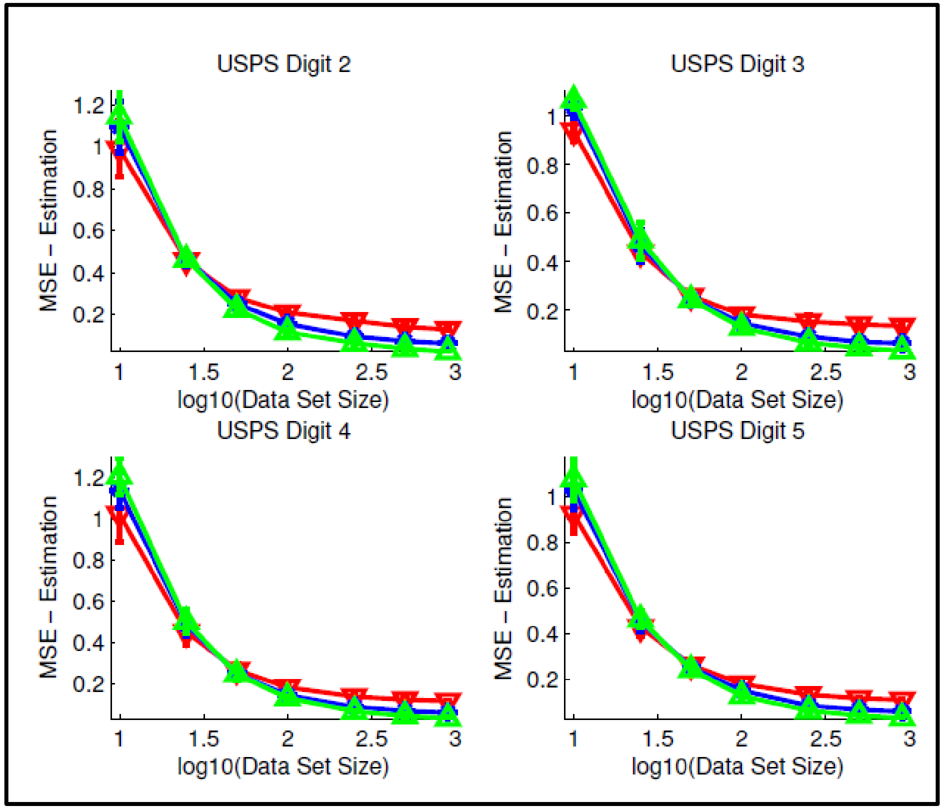
$$p(\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\theta}) \propto \exp\left(\sum_i \theta_i y_i + \sum_{ij} \theta_{ij} y_i y_j + \sum_i \theta_{ii} y_i x_i\right)$$



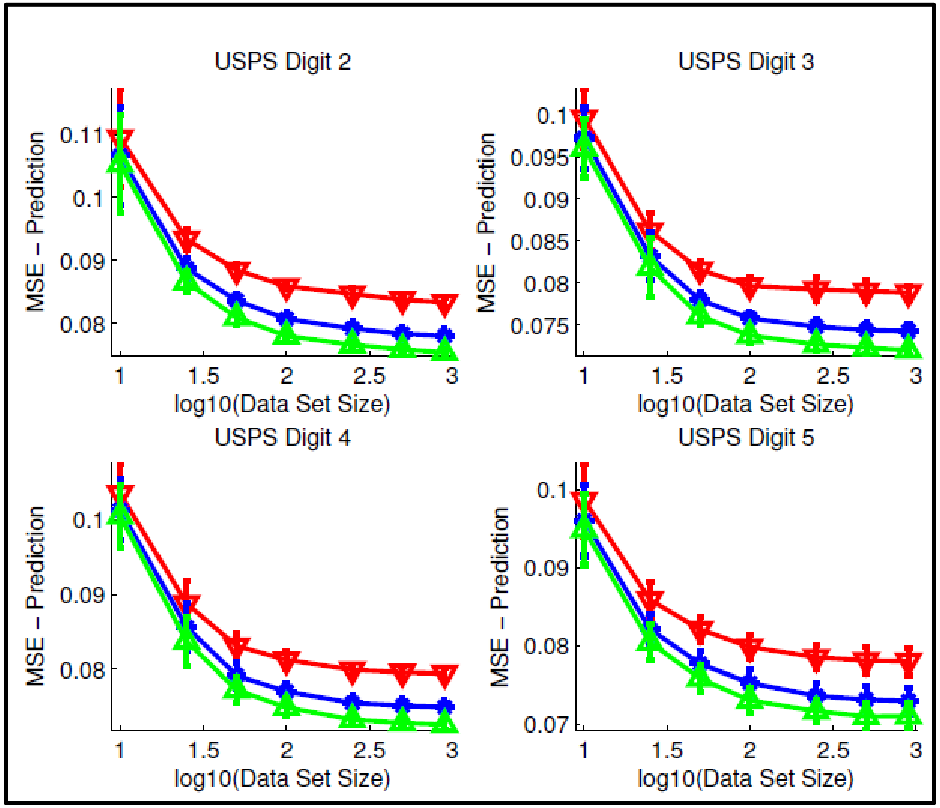
- Compute both Err_{est} and Err_{pred}
  - Err_{est} computed wrt $\ell_N(\boldsymbol{\theta})$ (using exact inference)

MRF De-Noising Experiments

Estimation Error



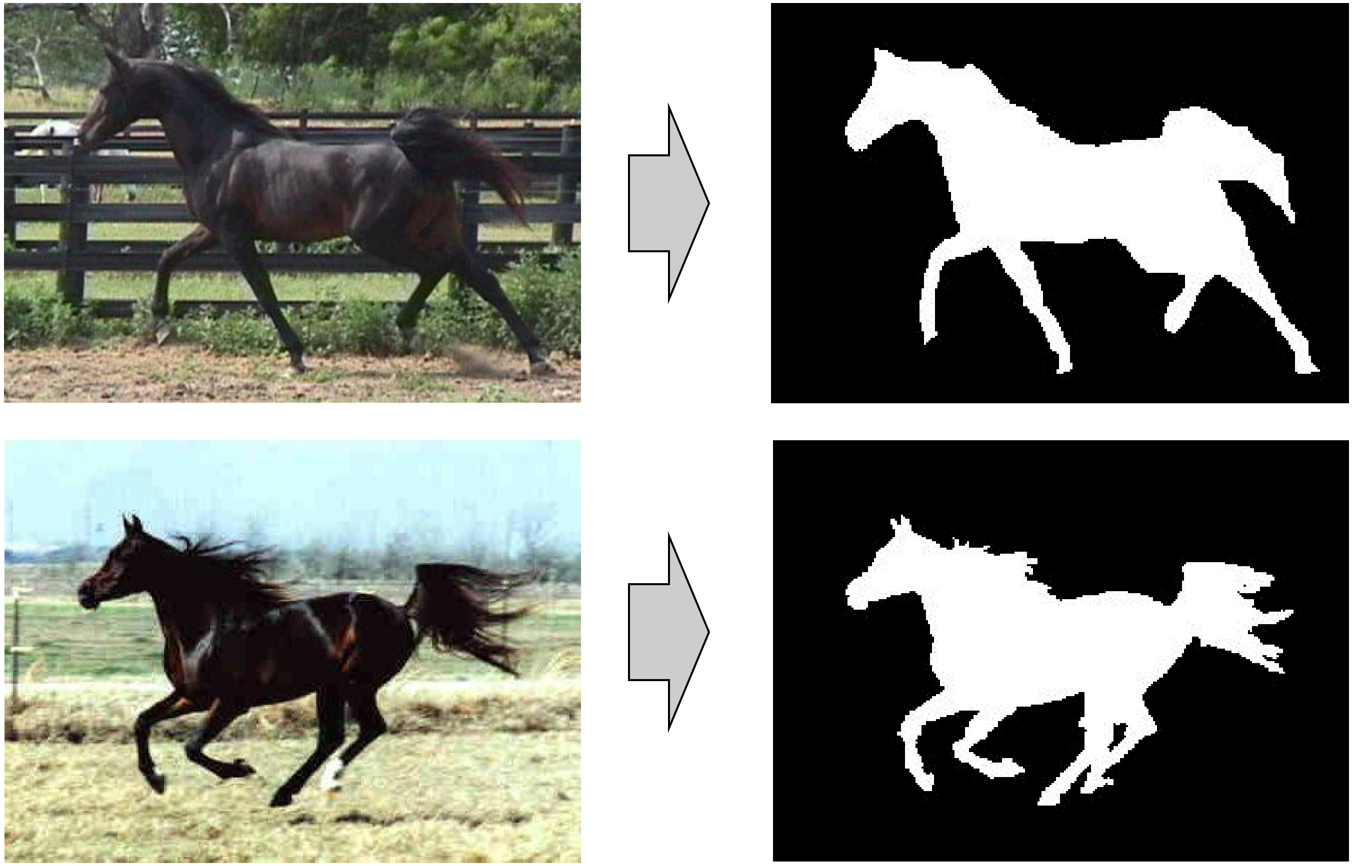
Prediction Error



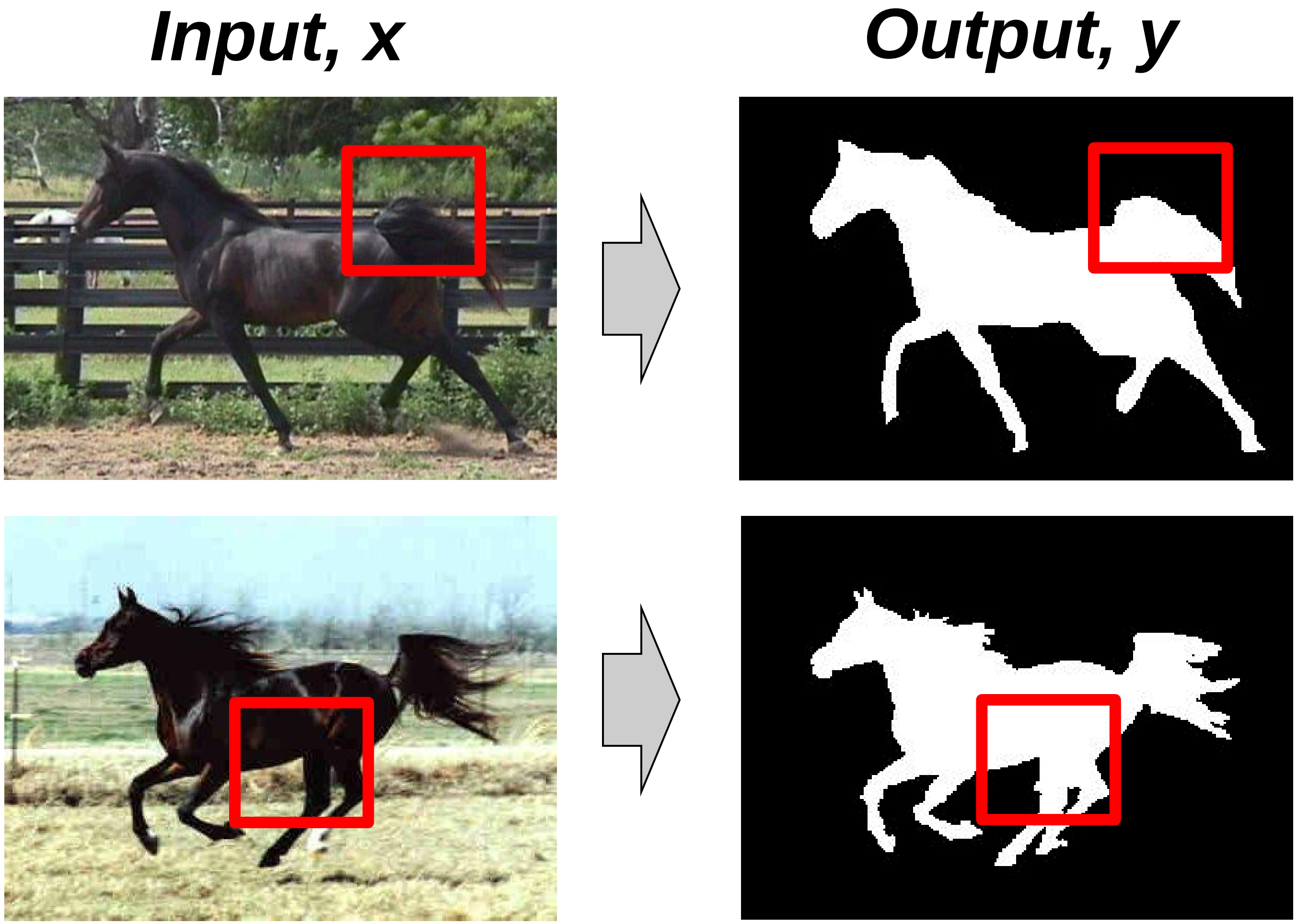
🟥 WMB$_2$ 🟦 WMB$_4$ 🟩 WMB$_8$

Foreground/Background Segmentation

Input, x ***Output, y***

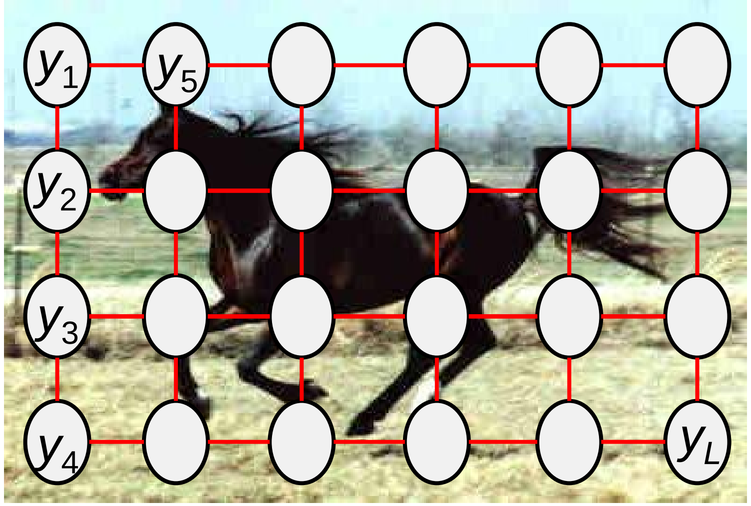


Foreground/Background Segmentation



Conditional Random Field Model

$$p(\mathbf{y}|\mathbf{x};\theta) \propto \exp\left(\sum_{i\in V}\sum\nolimits_u \theta_u\, f_u(y_i,\mathbf{x}) + \sum_{(i,j)\in E}\sum\nolimits_p \theta_p f_p(y_i,y_j,\mathbf{x})\right)$$



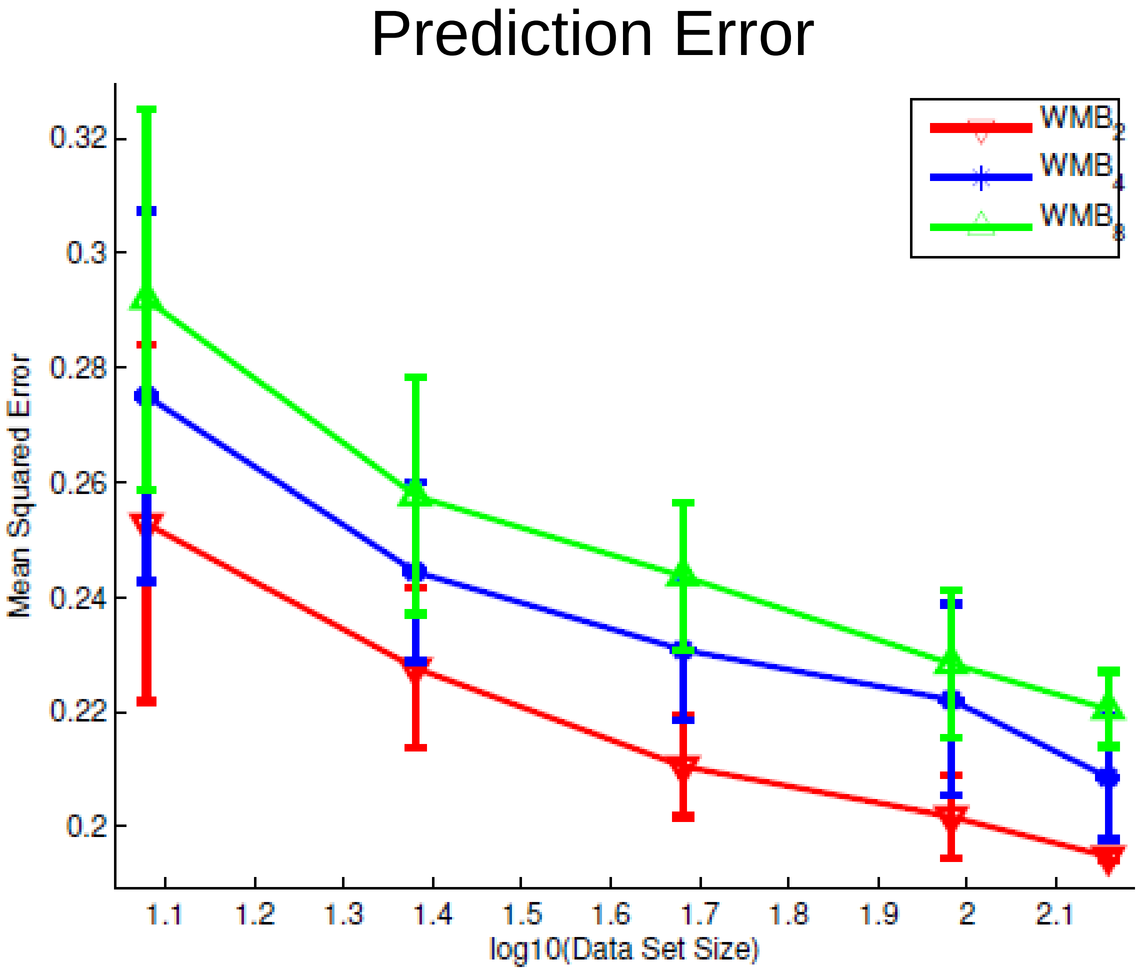
Unary Features:

$$f_u(y_i,x_i) = \begin{cases} \text{color}(x_i), & \text{if } y_i = \text{'horse'}. \\ 0, & \text{otherwise.} \end{cases}$$

Pairwise Features:

$$f_p(y_i,y_j,\mathbf{x}) = \begin{cases} 1, & \text{if } y_i = y_j \ , ||x_i - x_j|| < \epsilon \\ 0, & \text{otherwise.} \end{cases}$$

Quantitative Results

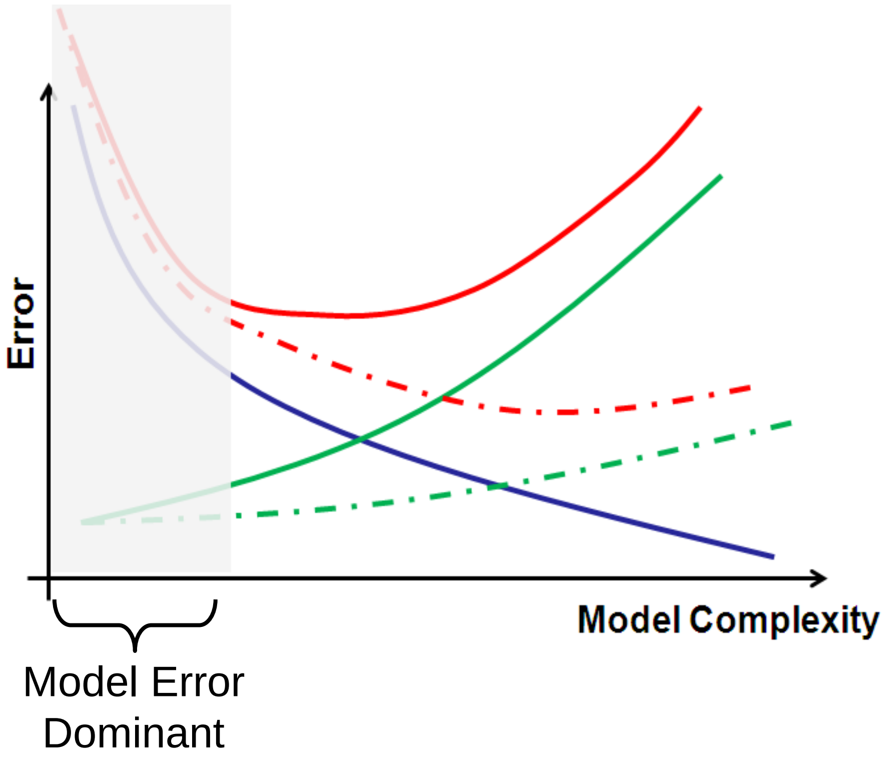


Qualitative Results

Summary of Experimental Study

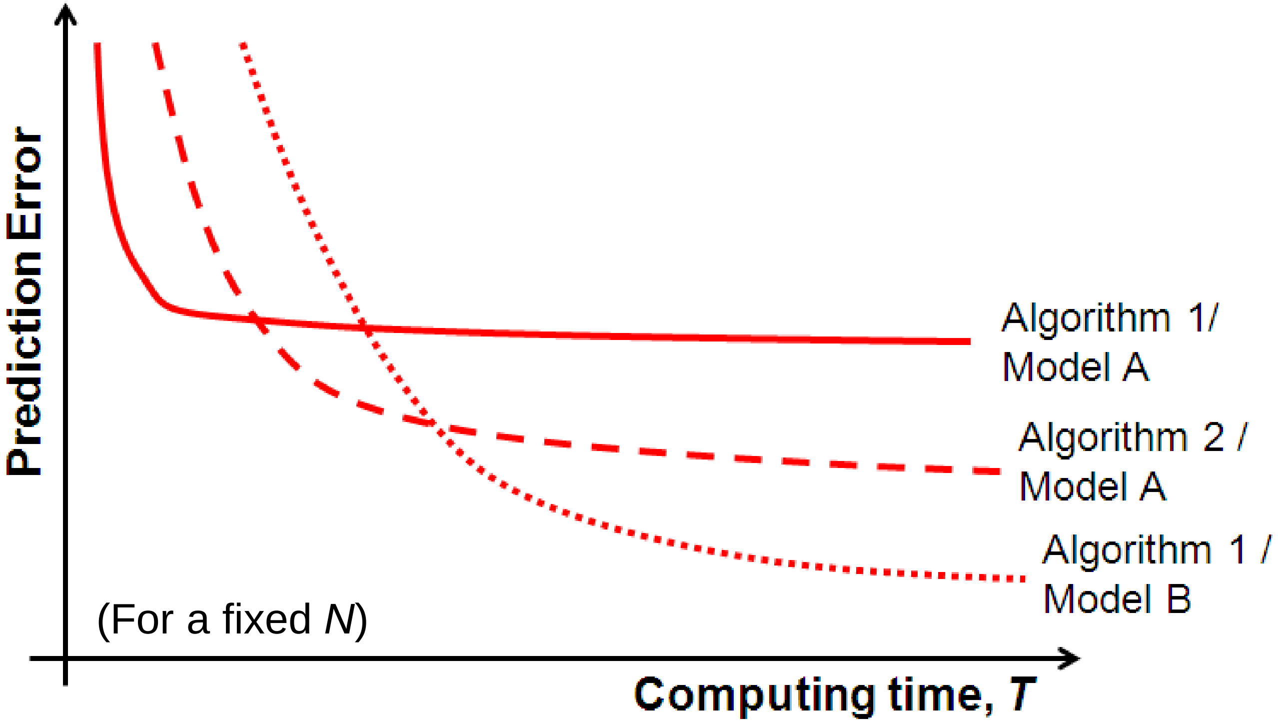
- ☐ If data set is small use a lower *iBound* method
  - Trade smaller variance for increased bias

- ☐ Higher *iBound* does not yield better predictions if model error is dominant



The Dream Scenario…

- ☐ Given data set *N* and time budget *T*, choose the model and algorithm that minimize test error



Outline of this Talk

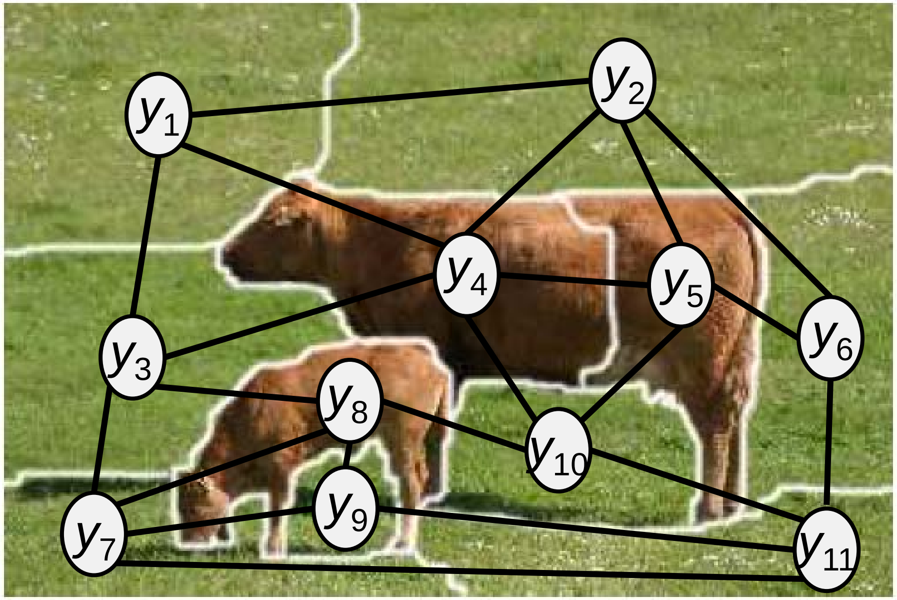
1. Max Likelihood Learning
    - Sources of error in likelihood-based learning
    - Computation-accuracy trade-offs in approximate learning

2. Computing Marginal Probabilities
    - Review of Belief Propagation (BP) & Generalized BP
    - Choosing Regions via Cycle Bases

3. Summary

Computing Marginals

After learning $p(\mathbf{y}) \propto \psi_{12}(y_1, y_2)\psi_{13}(y_1, y_3)\cdots$



…want to compute:

$p(y_2 = \text{'grass'})$?

$p(y_5 = \text{'cow'})$?

Variational Perspective [Wainwright & Jordan ‘08]

- Convert from a summation task…

$$\log Z = \log \sum_{\mathbf{y}\in\mathbf{Y}} \prod_{i\in V} \psi_i(y_i) \prod_{(i,j)\in E} \psi_{ij}(y_i, y_j)$$

…to an optimization task

$$\log Z = \underbrace{\max_{b\in\mathbb{P}}}_{\text{Set of valid probability distributions}} \left[E_b\left[\log\psi(\mathbf{y})\right] + \underbrace{H(\mathbf{y};b)]}_{\text{Entropy of distribution } b(\boldsymbol{y})}\right.$$

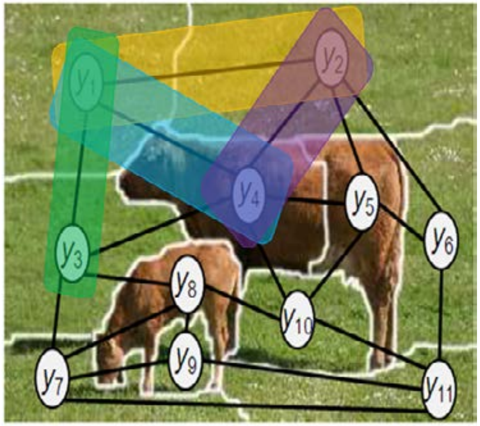
Variational Approximations

- GBP introduces *two* approximations

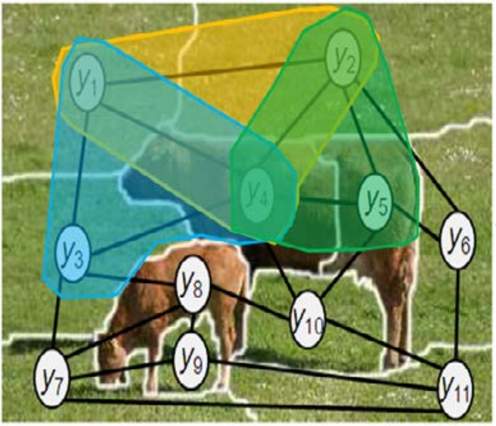
$$\log Z = \underbrace{\max_{b\in\mathbb{L}}}_{\text{1) Locally consistent beliefs}} \left[E_b \left[\log \psi(\mathbf{y})\right] + \underbrace{\tilde{H}(\mathbf{y};b)}_{\text{2) Approximate Entropy}} \right]$$

$$\tilde{H}_{\mathrm{GBP}}(\mathbf{y};b) = -\sum_{R\in\mathcal{R}} c_R \underbrace{H(\mathbf{y}_R;b)}_{\text{Marginal Entropy on Region } R \text{, where } \mathbf{y}_R \subset \mathbf{y}} = -\sum_{R\in\mathcal{R}} c_R \sum_{\mathbf{y}_R} b_R(\mathbf{y}_R) \log b_R(\mathbf{y}_R)$$

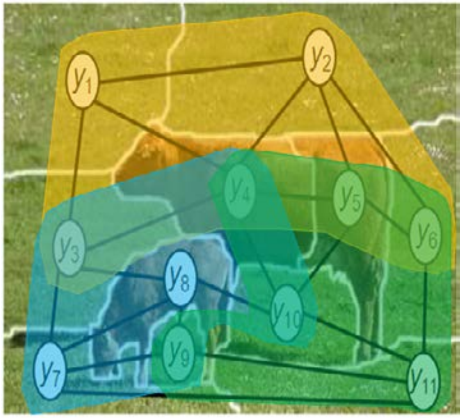
Which Regions do we choose?


Bethe / BP

●●●


Kikuchi / GBP

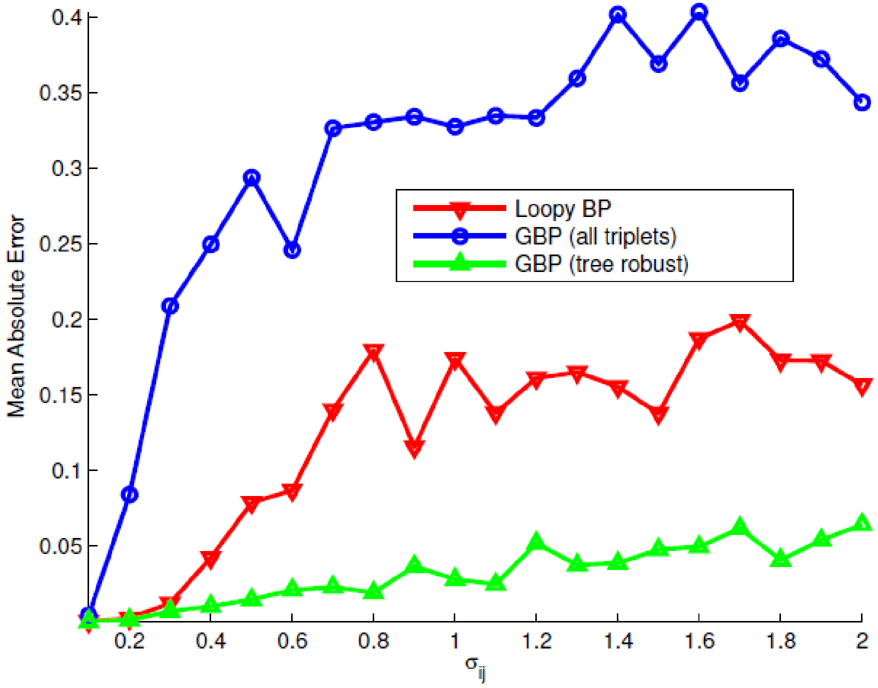
●●●


Exact

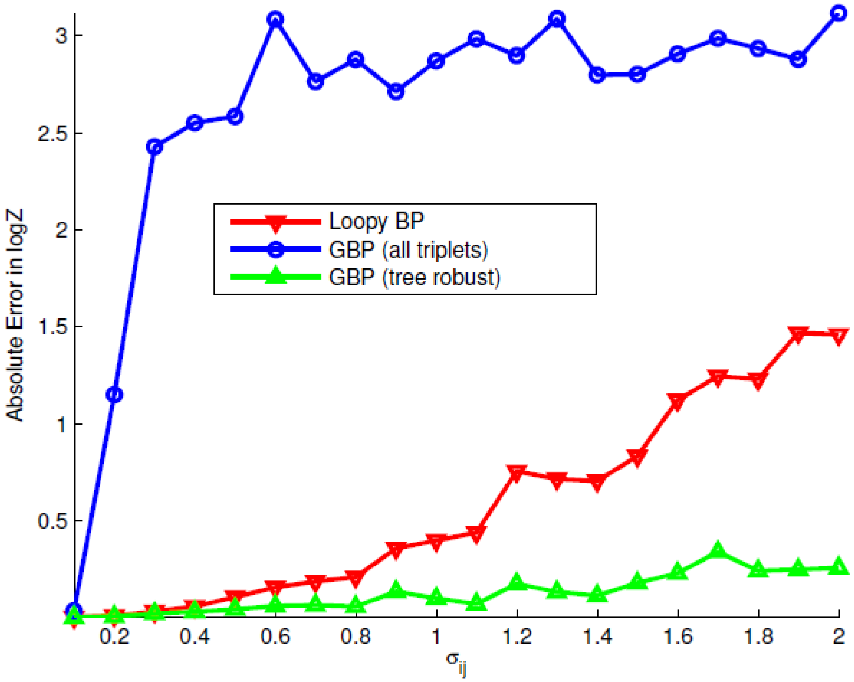
Complexity:	exp(2)	exp(|region|)	exp(w*)
Accuracy:	low	high	exact

Region choice is important!

Error in marginal estimates



Error in log Z estimates



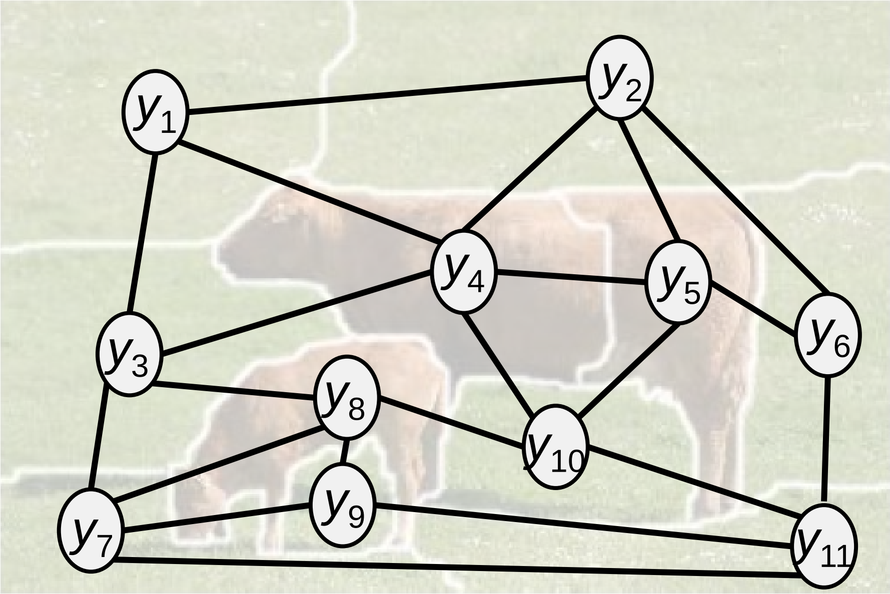
Same complexity; very different accuracies!

Existing Guidance on Region Choice

- Choose Regions so that:
  - $\tilde{H}_{\mathrm{GBP}}(\mathbf{y};b)$ exact when *p*(***y***) nearly uniform
    [Yedidia, Freeman, Weiss ’02] [Pakzad & Anantharam ‘05]
  - $\tilde{H}_{\mathrm{GBP}}(\mathbf{y};b)$ exact when *p*(***y***) nearly deterministic
    [Yedidia, Freeman, Weiss ’02]
  - All fixed points are uniform when *p*(***y***) is uniform
    [Welling, Minka, Teh ‘05]

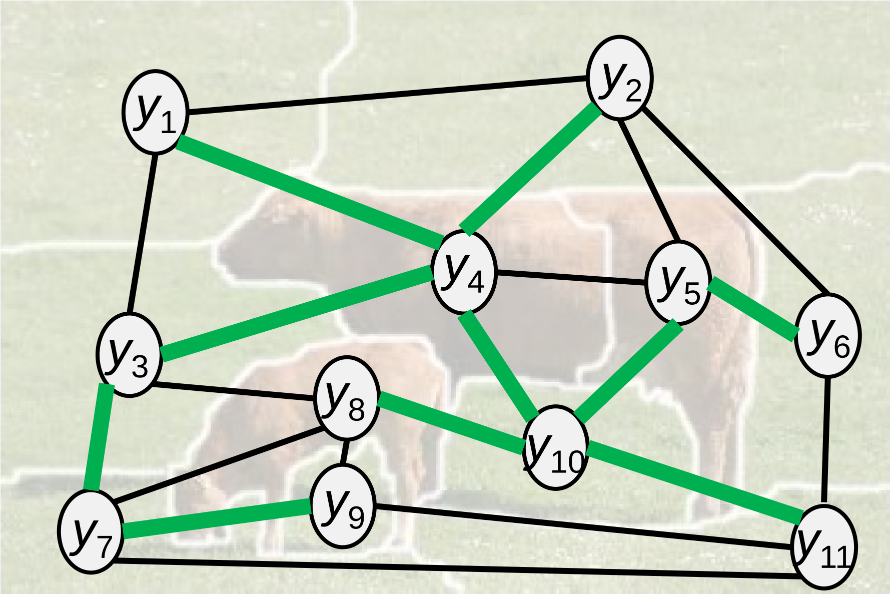
Tree-Robustness [Gelfand & Welling '12]

- ☐ Consider a pairwise model $p(\mathbf{y}) \propto \prod_{(i,j)\in E} \psi_{ij}(y_i, y_j)$



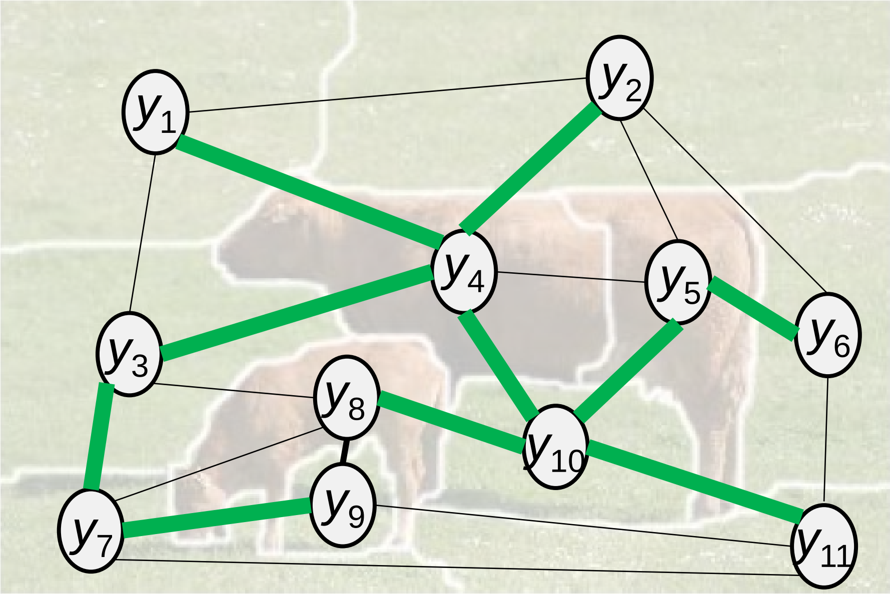
Tree-Robustness [Gelfand & Welling '12]

- ☐ Consider a pairwise model $p(\mathbf{y}) \propto \prod_{(i,j)\in E} \psi_{ij}(y_i, y_j)$
- ☐ Let $T \subseteq E$ be a tree in *G*



Tree-Robustness [Gelfand & Welling '12]

- Consider a pairwise model $p(\mathbf{y}) \propto \prod_{(i,j)\in E} \psi_{ij}(y_i, y_j)$
- Let $T \subseteq E$ be a tree in *G*

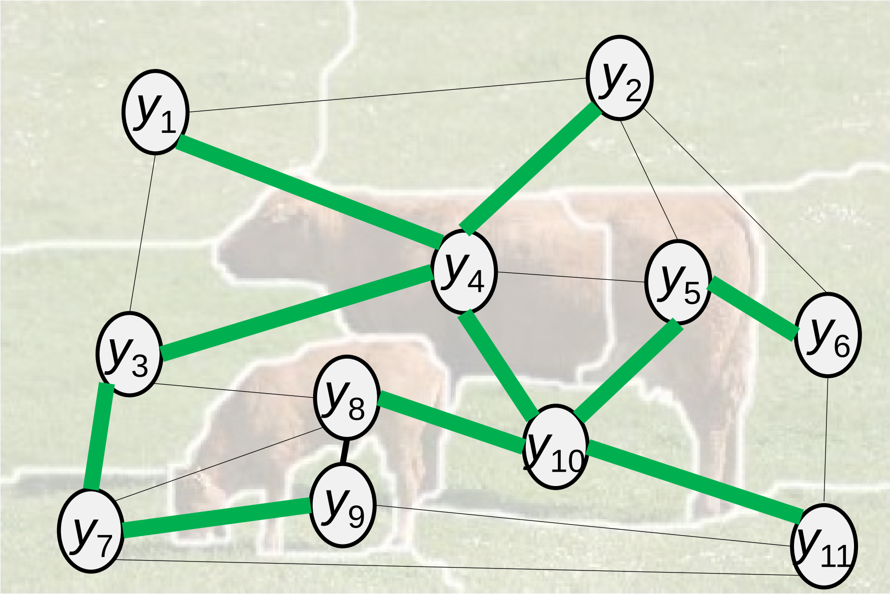


- Assume uniform off-tree factors:

$$\psi_{ij}(y_i, y_j) = \mathbf{1}$$

Tree-Robustness [Gelfand & Welling '12]

- Consider a pairwise model $p(\mathbf{y}) \propto \prod_{(i,j) \in E} \psi_{ij}(y_i, y_j)$
- Let $T \subseteq E$ be a tree in *G*



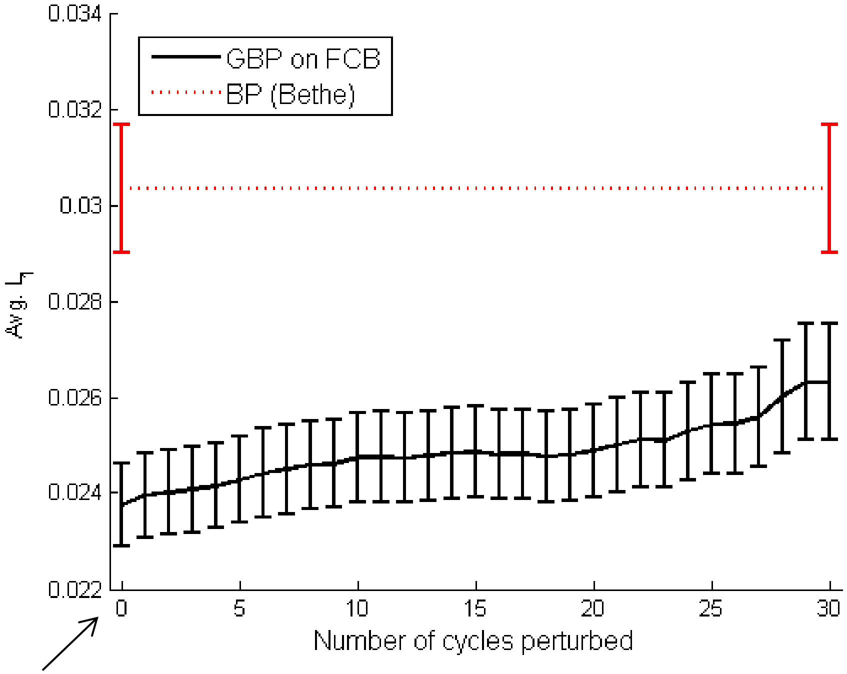
- Assume uniform off-tree factors:

$$\psi_{ij}(y_i, y_j) = \mathbf{1}$$

- $\tilde{H}_{\mathrm{GBP}}(\mathbf{y}; b)$ is exact on $p_T(\mathbf{y})$ and all such trees in *G*!

Is Tree Robustness Desirable?

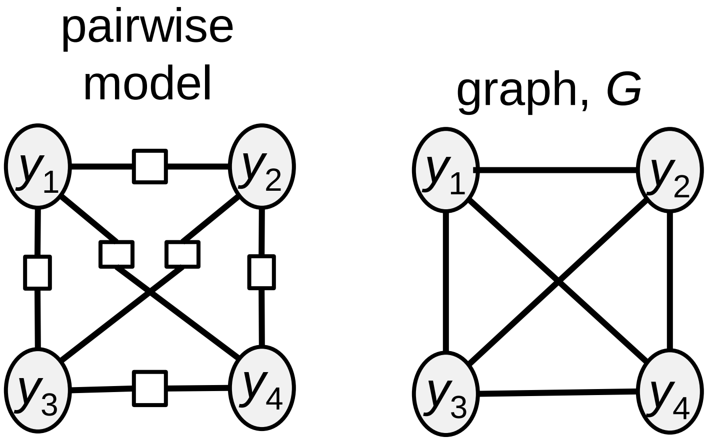
Accuracy degrades as Regions becomes less Tree Robust!



Tree Robust → Less Tree Robust

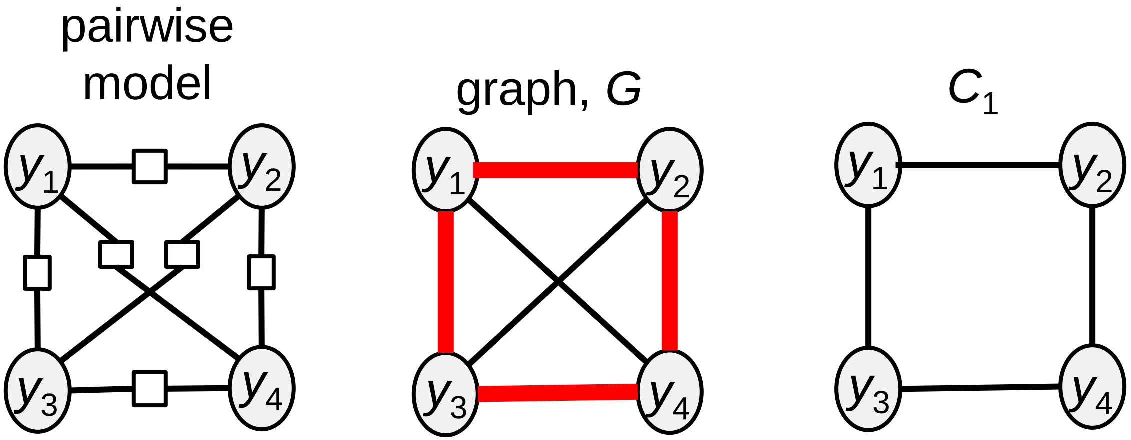
Bottom-Up Region Selection

- ☐ Selecting Regions ≡ Finding Cycle Bases in *G*



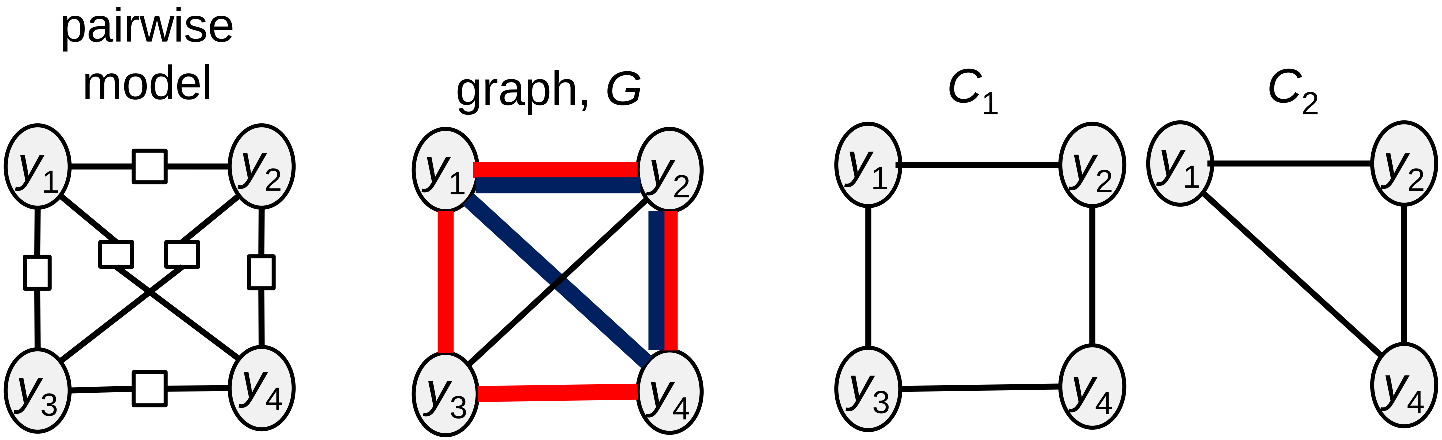
Bottom-Up Region Selection

- ❑ Selecting Regions ≡ Finding Cycle Bases in *G*



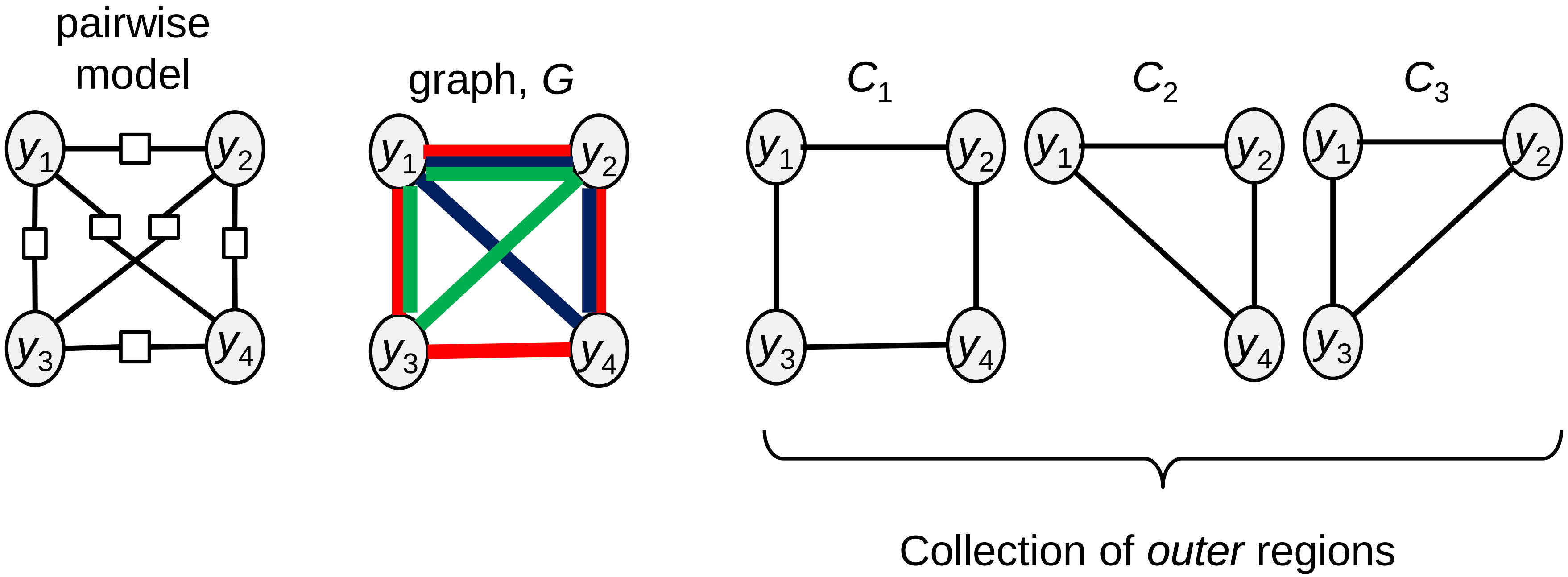
Bottom-Up Region Selection

- ☐ Selecting Regions ≡ Finding Cycle Bases in *G*



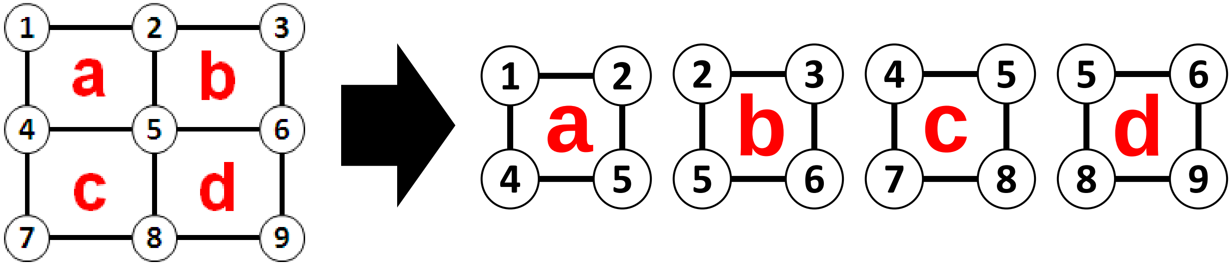
Bottom-Up Region Selection

- ☐ Selecting Regions ≡ Finding Cycle Bases in *G*

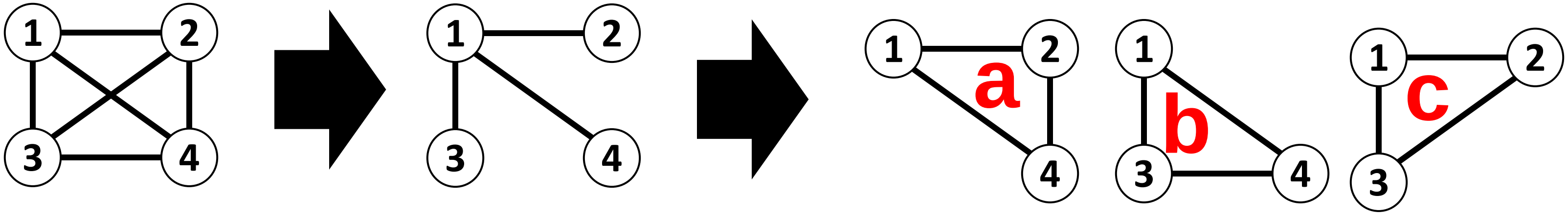


Identifying Tree-Robust Regions

- Selecting TR Regions ≡ Finding TR Cycle Basis
    - Faces of a planar graph:

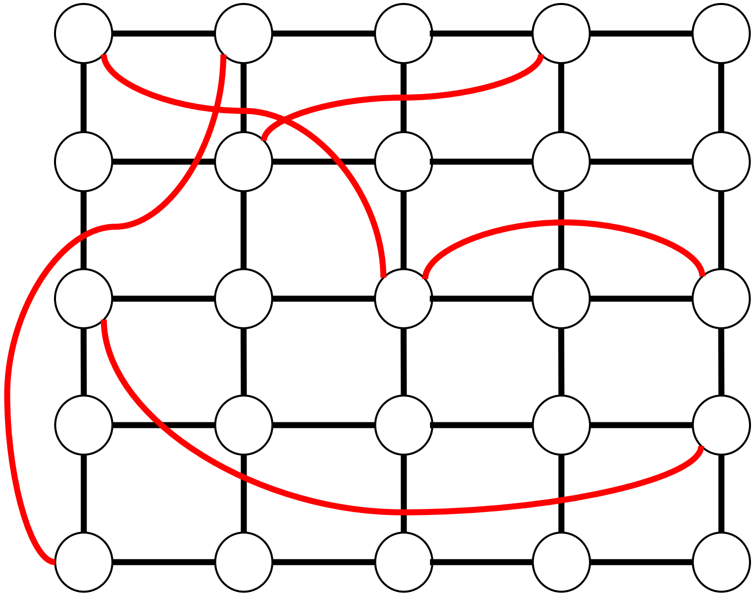


    - ‘Star’ Construction



Experimental Results

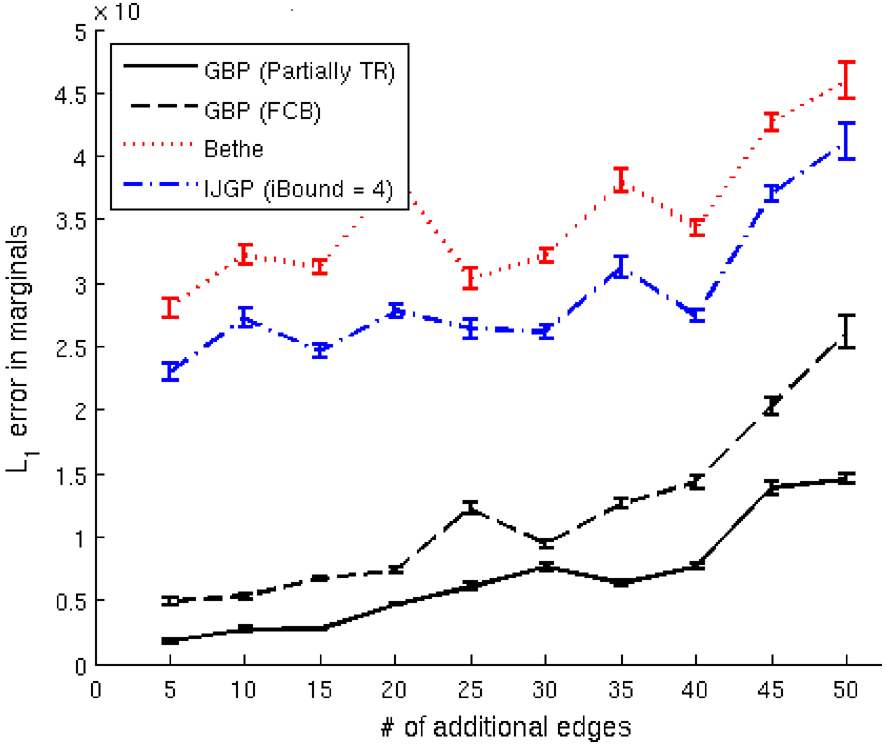
- ☐ Grids with long range interactions

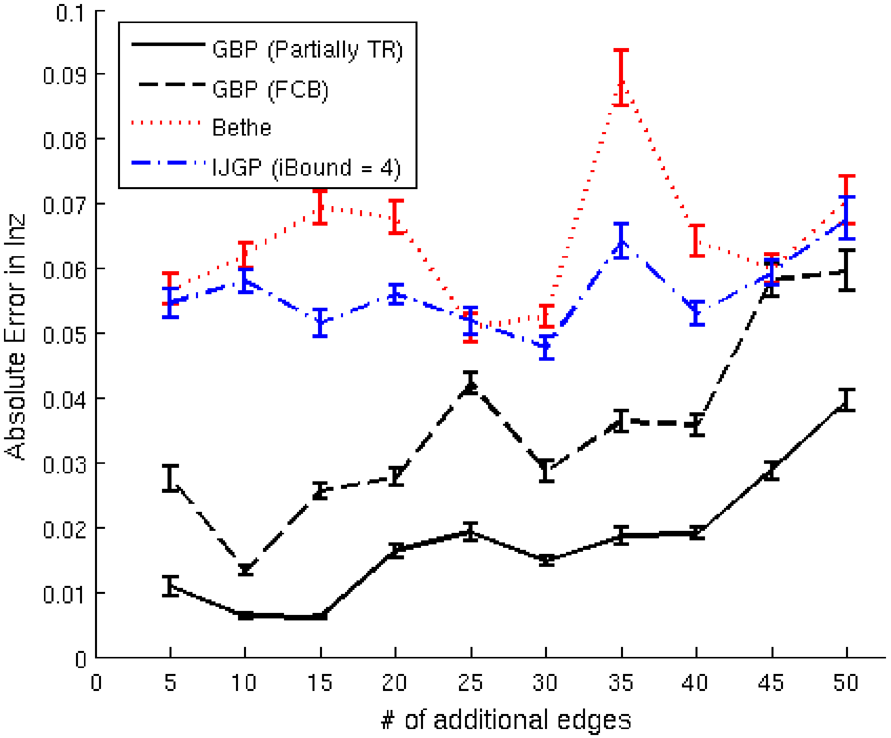


Algorithms

1. GBP with Tree-Robust Core
2. GBP with 'ear' construction
3. Loopy BP
4. Iterative Join Graph Prop. (IJGP) w/ *iBound* = 4

Experimental Results


Error in marginal estimates


Error in log Z estimates

→ Increasing # of non-planar edges

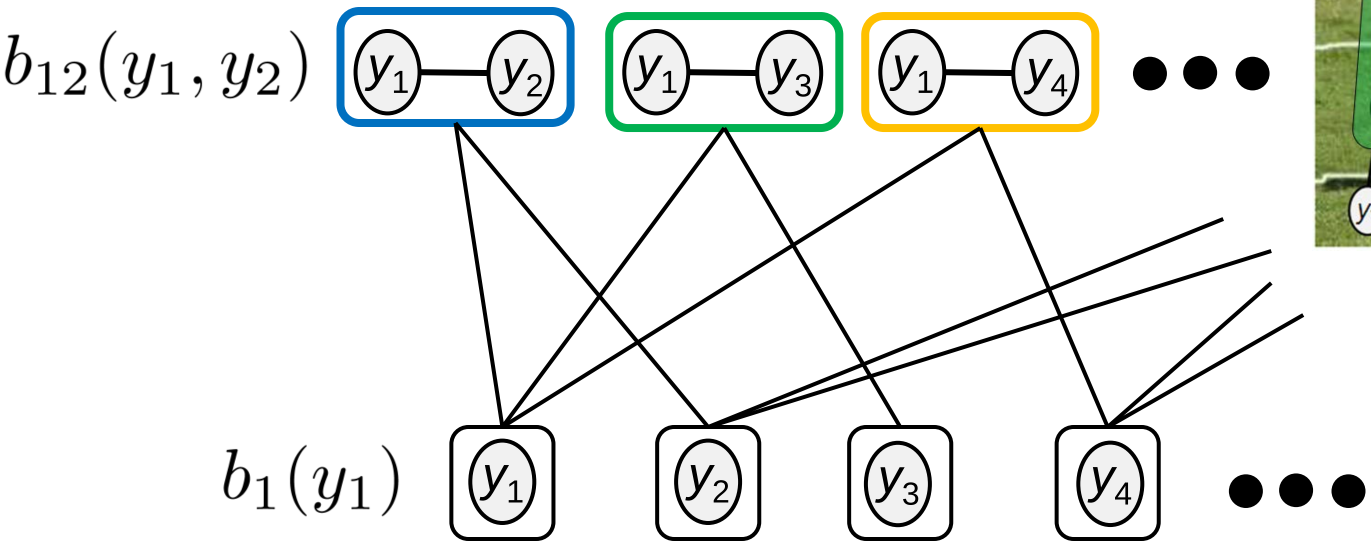
Outline of this Talk

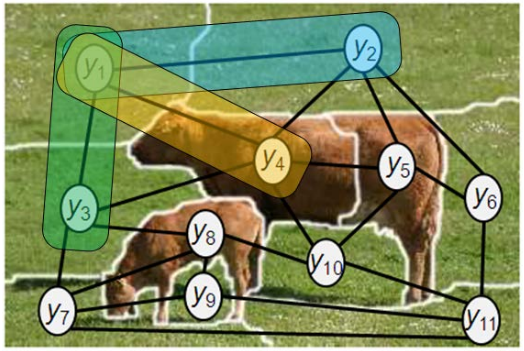
1. Max Likelihood Learning
   - Sources of error in likelihood-based learning
   - Computation-accuracy trade-offs in approximate learning
2. Computing Marginal Probabilities
   - Review of Belief Propagation (BP) & Generalized BP
   - Choosing Regions via Cycle Bases
3. Summary

Summary

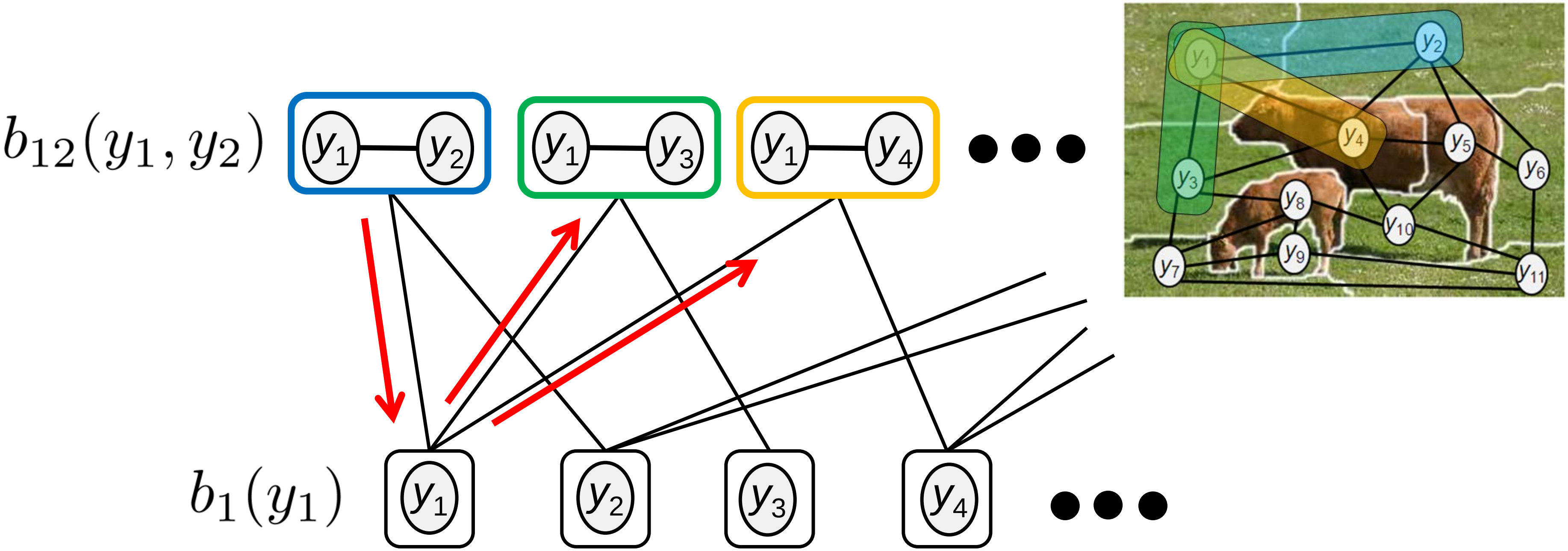
- Continued need for *better* inference methods
  - Advocate a “bottom-up” approach to inference
- Computation-Accuracy Trade-offs in Learning
  - Small N => Low computation inference
  - *Better* inference does not mean *better* predictions
- Proposed tree-robustness for choosing regions
- Connected finding regions to finding cycle bases

Belief Propagation (BP) [Pearl '88]





Belief Propagation (BP) [Pearl '88]

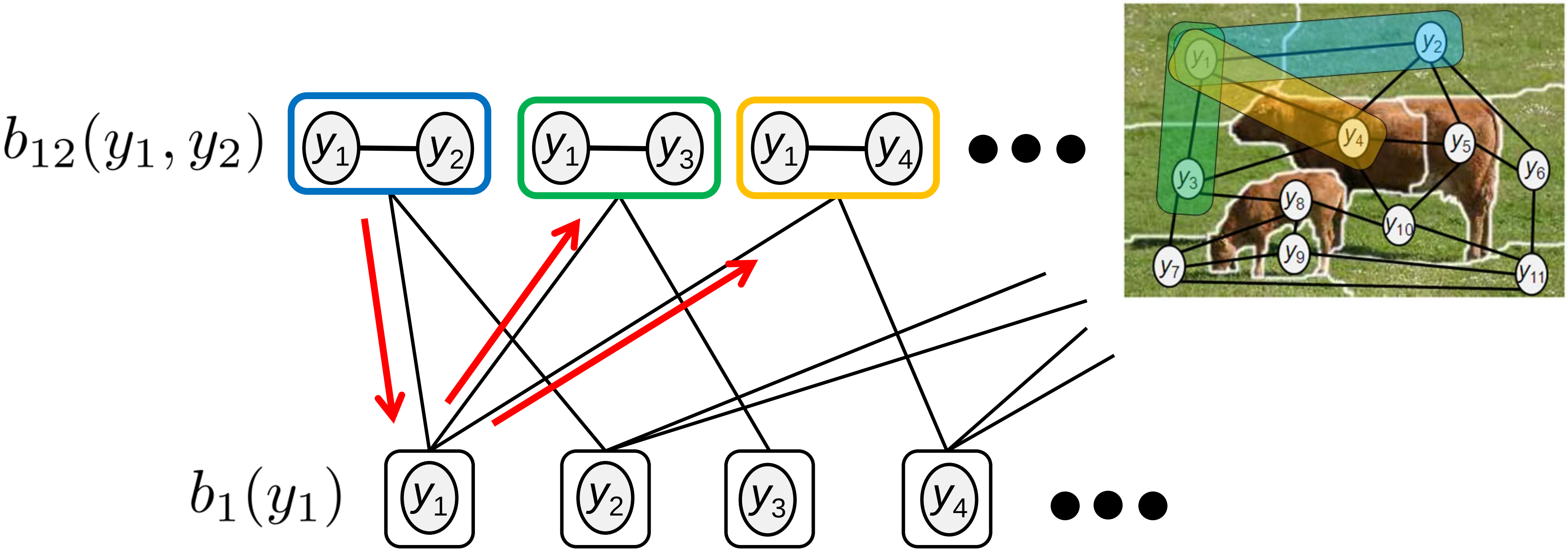


$$m_{12\to1}(y_1) = \frac{\sum_{y_2} b_{12}(y_1, y_2)}{b_1(y_1)} \qquad b_1(y_1) \leftarrow b_1(y_1) m_{12\to1}(y_1)$$

$$= m_{1\to13}(y_1)$$

$$= m_{1\to14}(y_1)$$

Belief Propagation (BP) [Pearl '88]

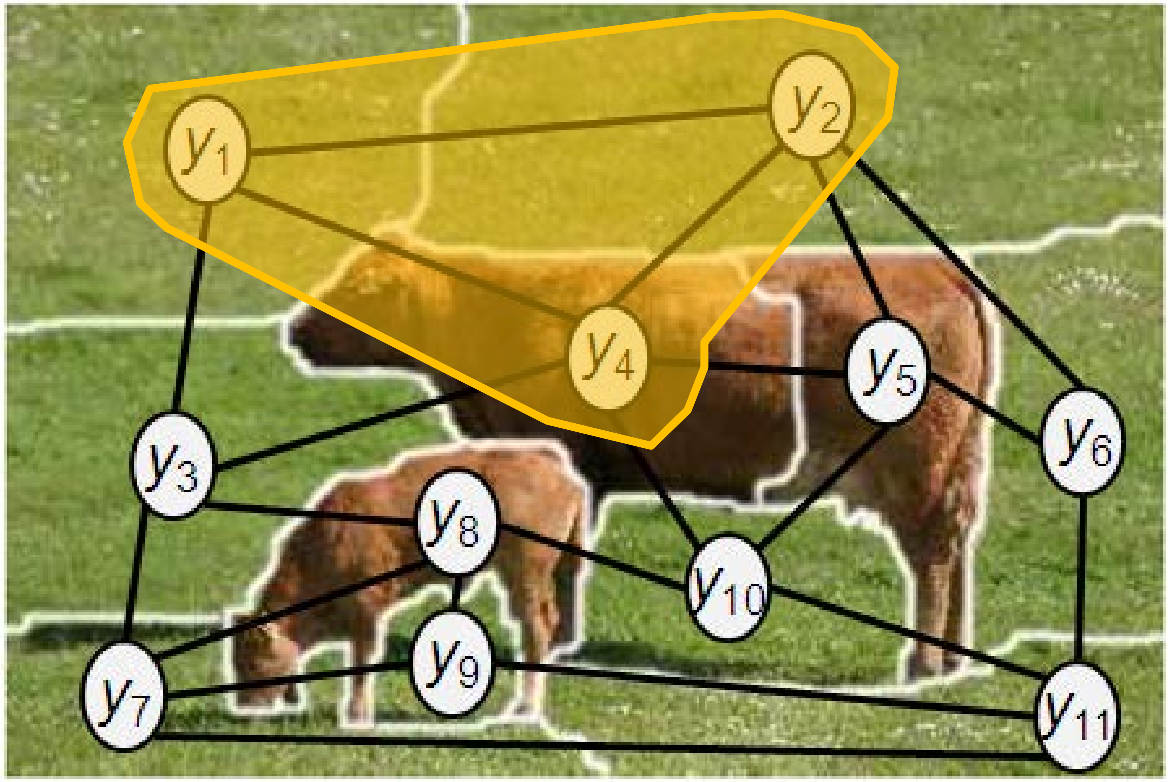


$b_{12}(y_1, y_2)$

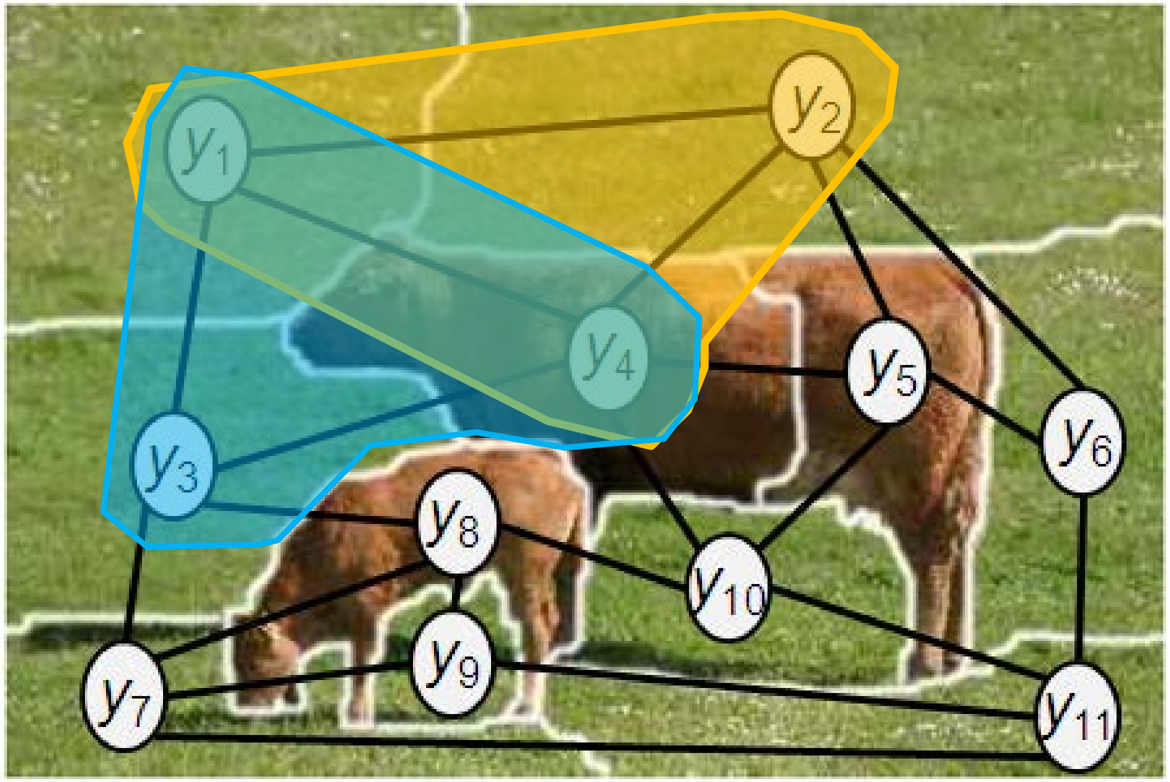
$b_1(y_1)$

Iterate until beliefs are *consistent* !

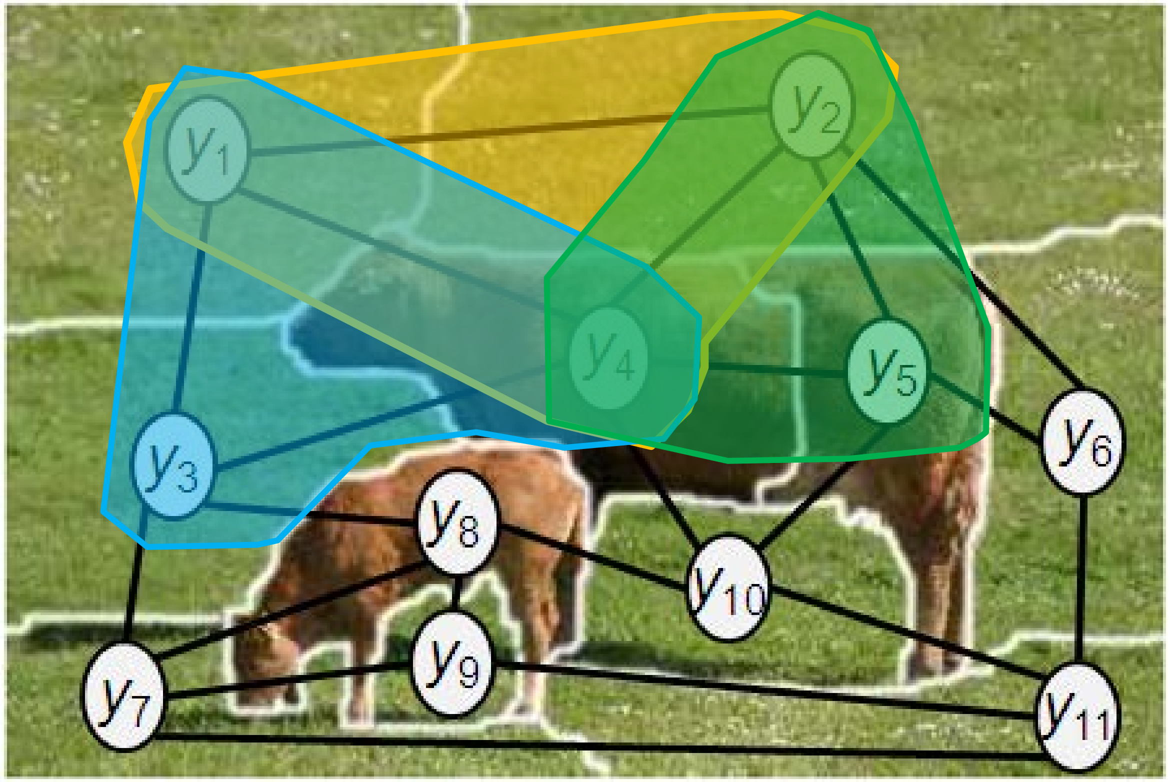
Generalized BP [Yedidia, Freeman, Weiss '00]



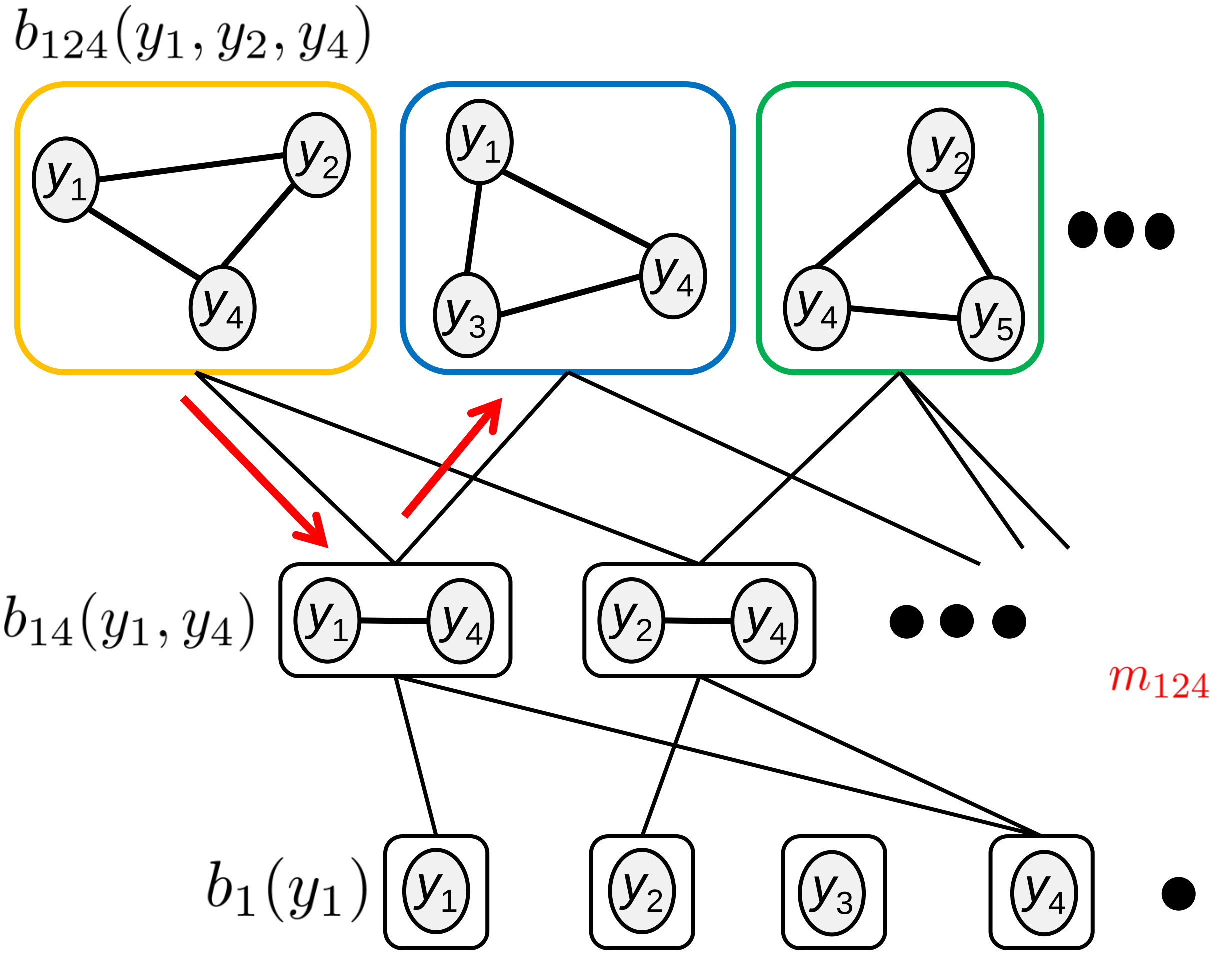
Generalized BP [Yedidia, Freeman, Weiss ‘00]

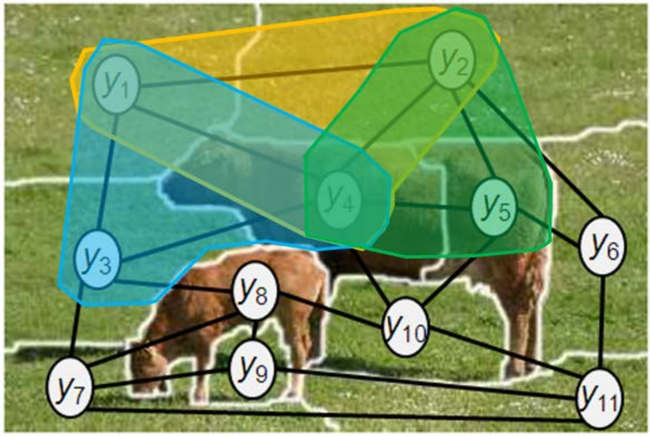


Generalized BP [Yedidia, Freeman, Weiss ‘00]



Generalized BP [Yedidia, Freeman, Weiss ‘00]

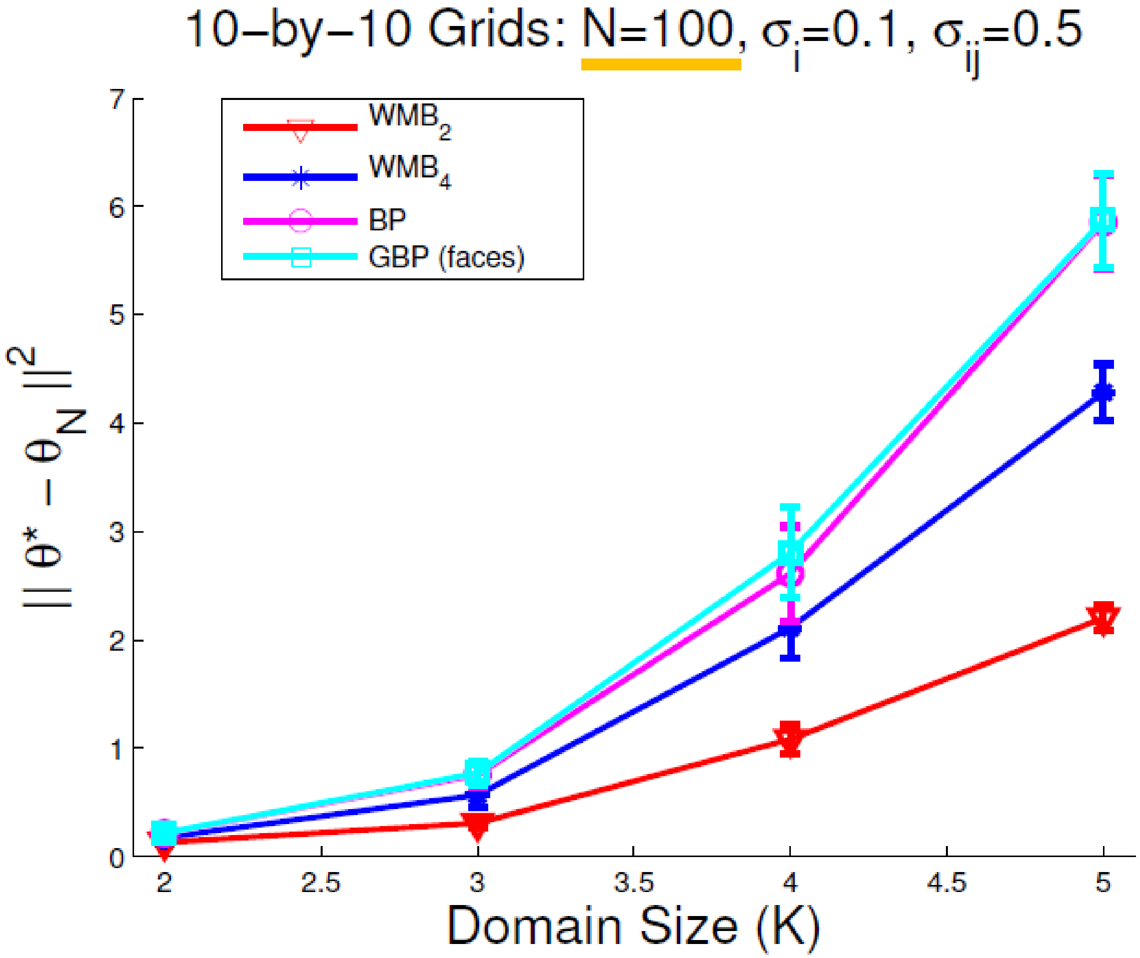




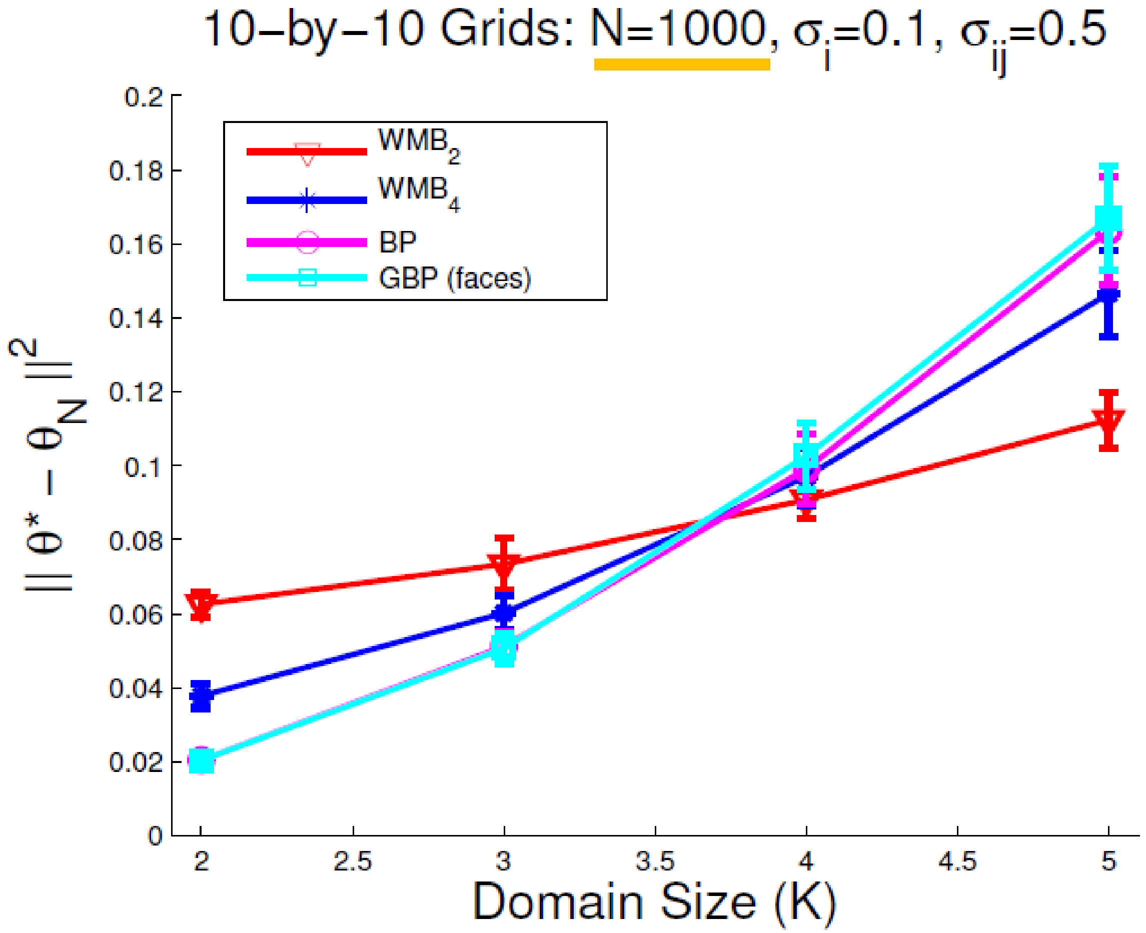
$$m_{124\to 14}(y_1, y_4) = \frac{\sum_{y_2} b_{124}(y_1,y_2,y_4)}{b_{14}(y_1,y_4)}$$

$$= m_{14\to 134}(y_1, y_4)$$

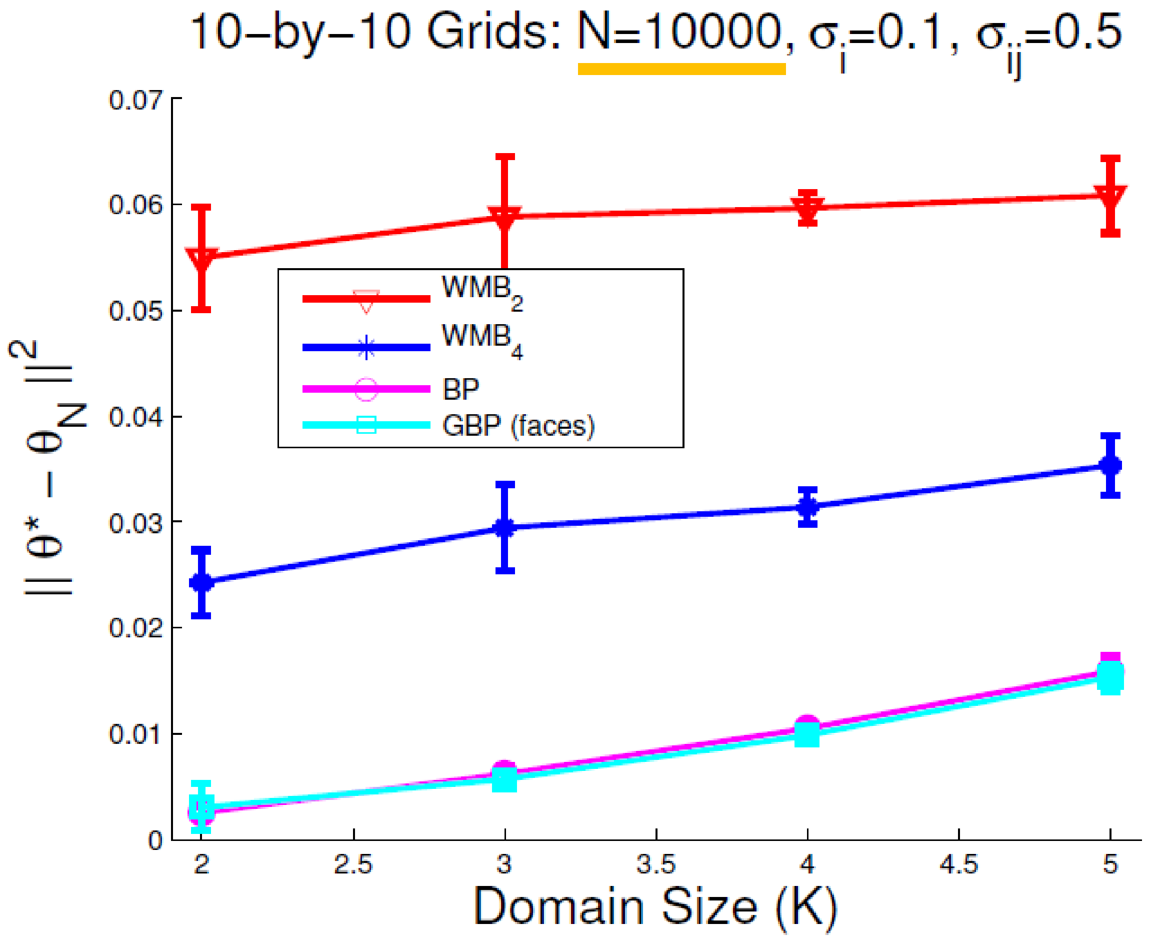
Domain Size Experiments



Domain Size Experiments



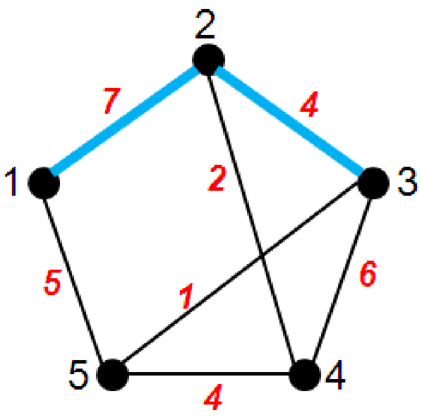
Domain Size Experiments



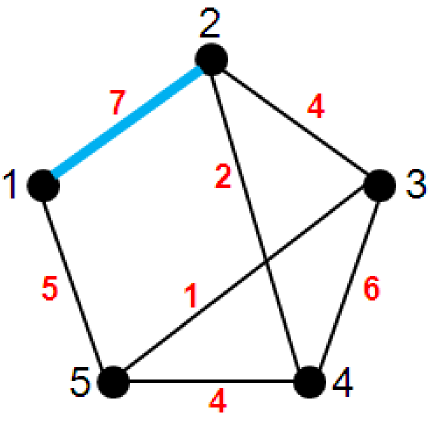
Weighted Matching Problem

- ☐ Given graph *G*=(*V*,*E*) with edge weights $\{w_e\}_{e \in E}$ find a *matching* of total maximum weight
  - Matching: subset of *E*, such that no 2 edges share a vertex

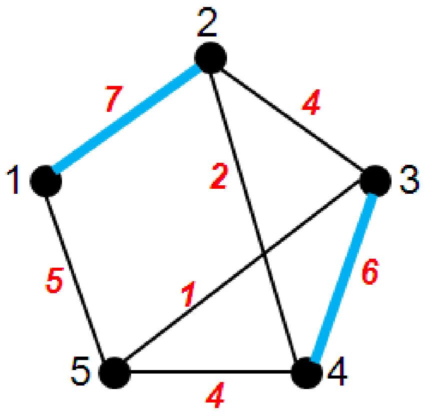
Not a Matching



Valid Matching



Max Weight Matching



Solving Weighted Matching Problems

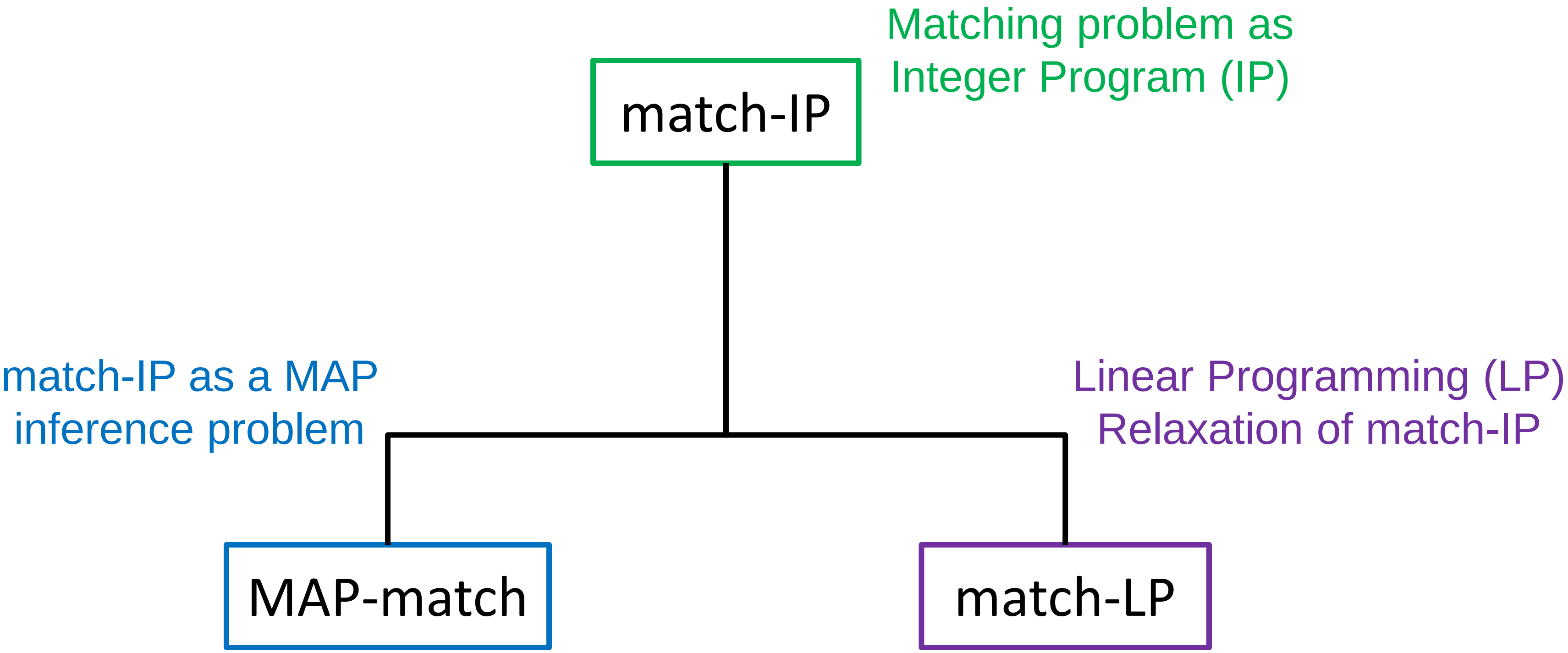
- Many efficient algorithms exist

Edmonds	1965	$O(\|V\|^2 \|E\|)$
Lawler	1973	$O(\|V\|^3)$
Gabow	1974	$O(\|V\|^3)$
Galil, Micali, Gabow	1986	$O(\|V\|\|E\| \log\|V\|)$
Gabow	1990	$O(\|V\|(\|E\|+\|V\|\log\|V\|))$

- Recent results using Belief Propagation

Bayati, Shah, Sharma [bipartite graphs]	2005	$O(c \|V\|^3)$
Sanghavi,Malioutov, Willsky [general graphs]	2007	$O(c\, w^{max}\|V\|^3)$

Solving Weighted Matching Problems



Matching problem as Integer Program (IP)

match-IP

match-IP as a MAP inference problem

Linear Programming (LP) Relaxation of match-IP

MAP-match

match-LP

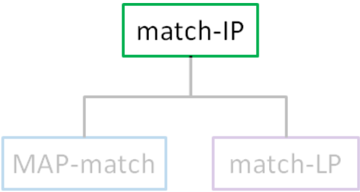
Matching as an Integer Program (IP)

❑ Given graph *G*=(*V*,*E*) with edge weights $\{w_e\}_{e\in E}$

$$\text{match-IP}: \quad \max_{\mathbf{x}} \textstyle\sum_e w_e x_e$$

$$\text{s.t.} \quad \underbrace{\underbrace{\sum_{e\in\delta(i)} x_e \le 1 \ \ \forall i \in V,}_{\text{Vertex Incidence Constraint}} \quad \underbrace{x_e \in \{0,1\} \ \ \forall e \in E}_{\text{Integrality Constraint}}}$$

Together define *matching polytope*, $\mathcal{P}_{\text{match-IP}}(G)$



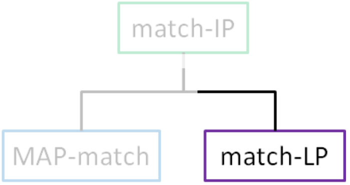
Linear Programming (LP) Relaxation

- Given graph *G*=(*V*,*E*) with edge weights $\{w_e\}_{e\in E}$

$$\text{match-}{\color{red}\text{LP}}: \quad \max_{\mathbf{x}} \textstyle\sum_e w_e x_e$$

$$\text{s.t.} \quad \underbrace{\underbrace{\sum_{e\in\delta(i)} x_e \le 1 \;\; \forall i \in V,}_{\text{Vertex Incidence Constraint}} \quad \underbrace{x_e \in {\color{red}[0,1]} \;\; \forall e \in E}_{\textit{Relaxed}\text{ Integrality Constraint}}}$$

Define a relaxed *matching polytope*, $\mathcal{P}_{\text{match-LP}}(G)$



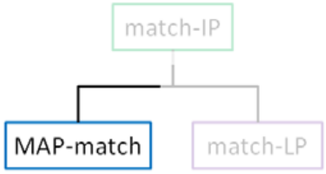
Matching as MAP Inference

- ☐ Associate a variable with each edge $\mathbf{x} = \{x_e\}_{e \in E}$

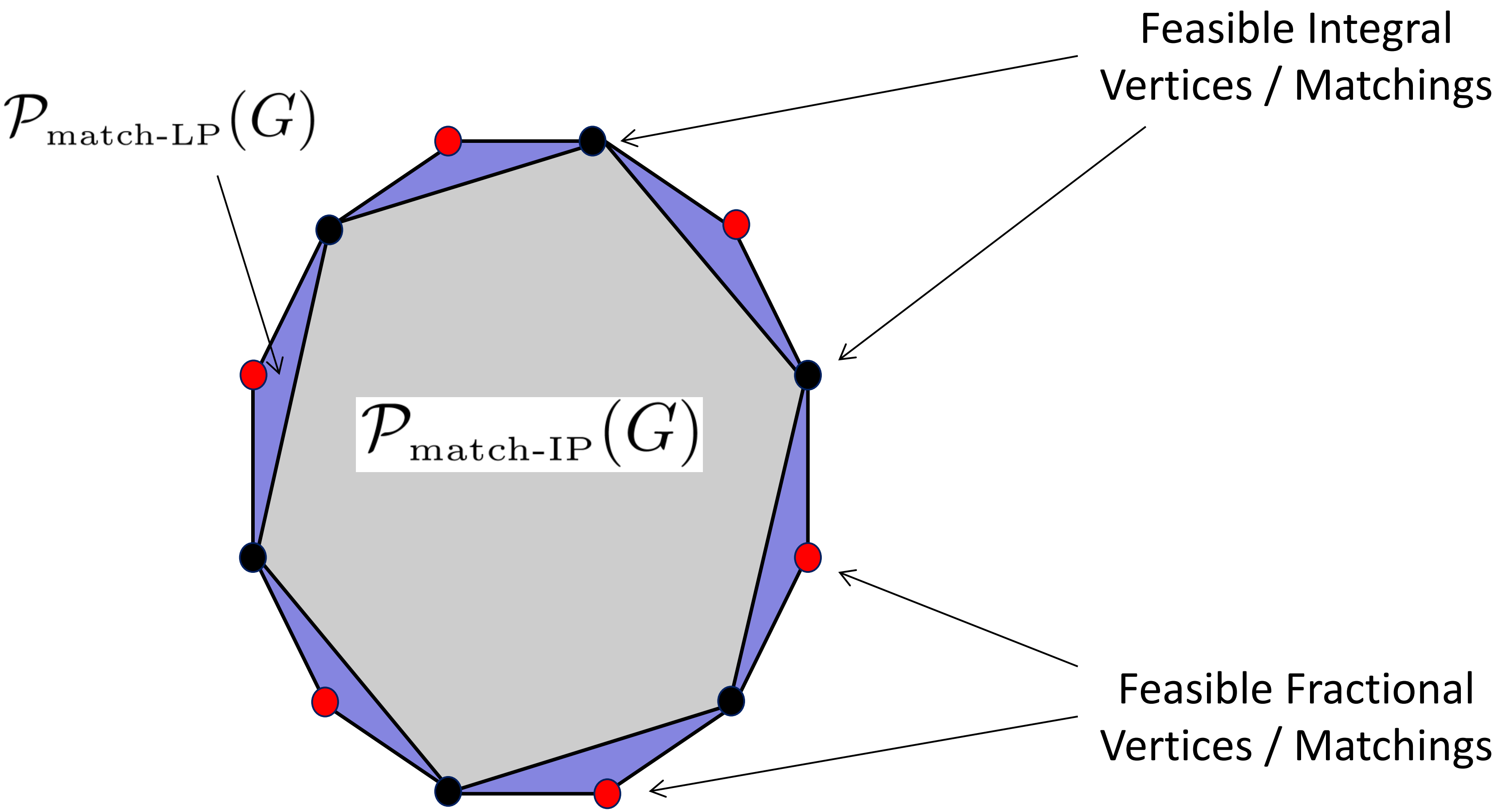
$$\text{match-MAP}: \quad \max_{\mathbf{x}} \sum_{e \in E} \theta_e(x_e) + \sum_{i \in V} \theta_i(\mathbf{x}_i)$$

where

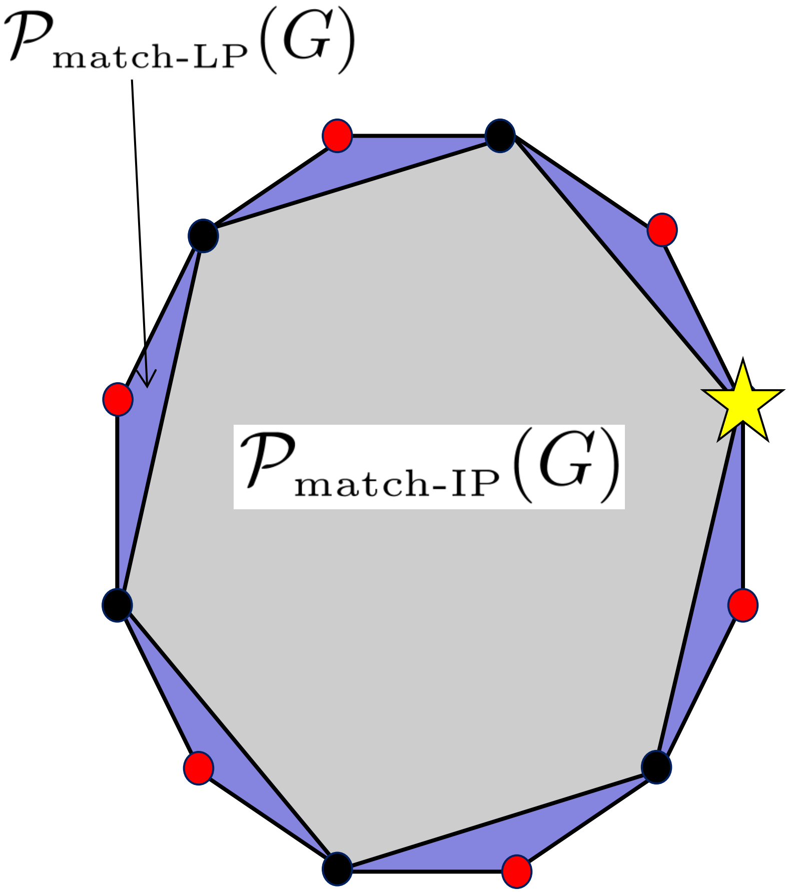
$$\underbrace{\theta_e(x_e) = w_e x_e}_{\text{Edge weights}} \qquad \underbrace{\theta_i(\mathbf{x}_i) = \begin{cases} 0 & \text{if } \sum_{e \in \delta(i)} x_e \leq 1 \\ -\infty & \text{otherwise} \end{cases}}_{\text{Vertex incidence constraints}}$$



Relationship between polytopes

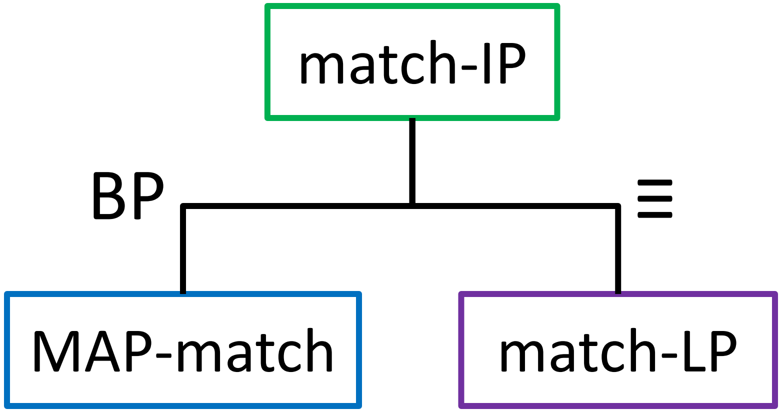


Relationship between polytopes



If integral solution to match-LP, **then** max-product BP on MAP-match is provably exact

bipartite graphs [Bayati, Shah, Sharma ‘05]
general graphs [Sanghavi, Malioutov, Willsky ‘07]



match-IP

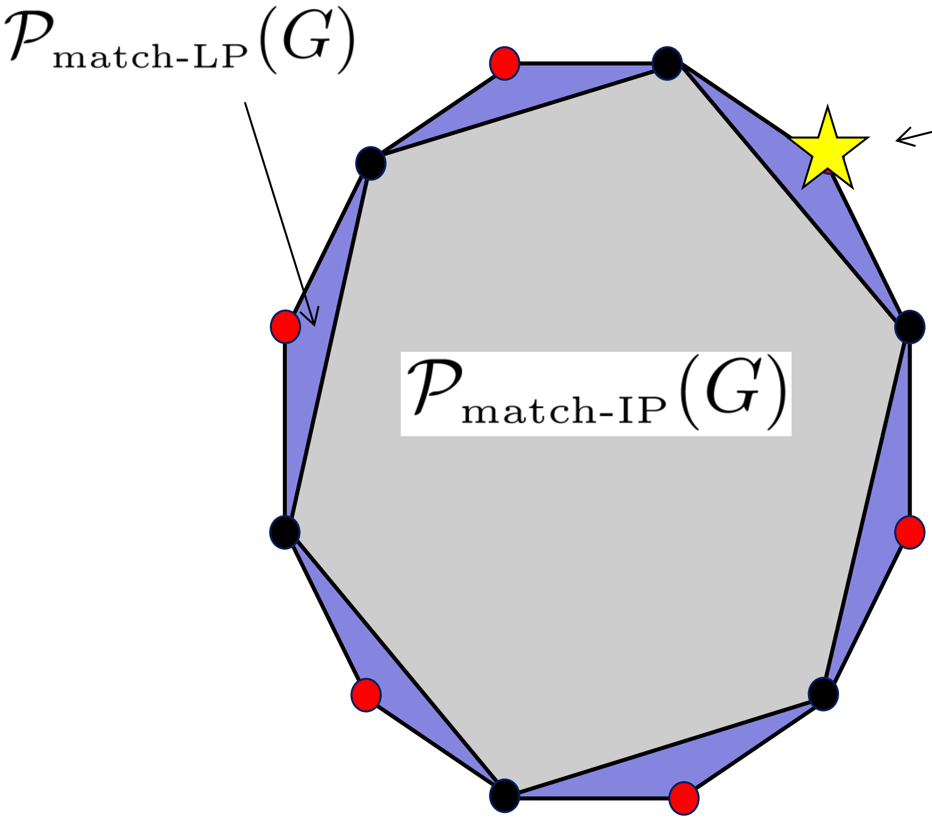
BP

≡

MAP-match

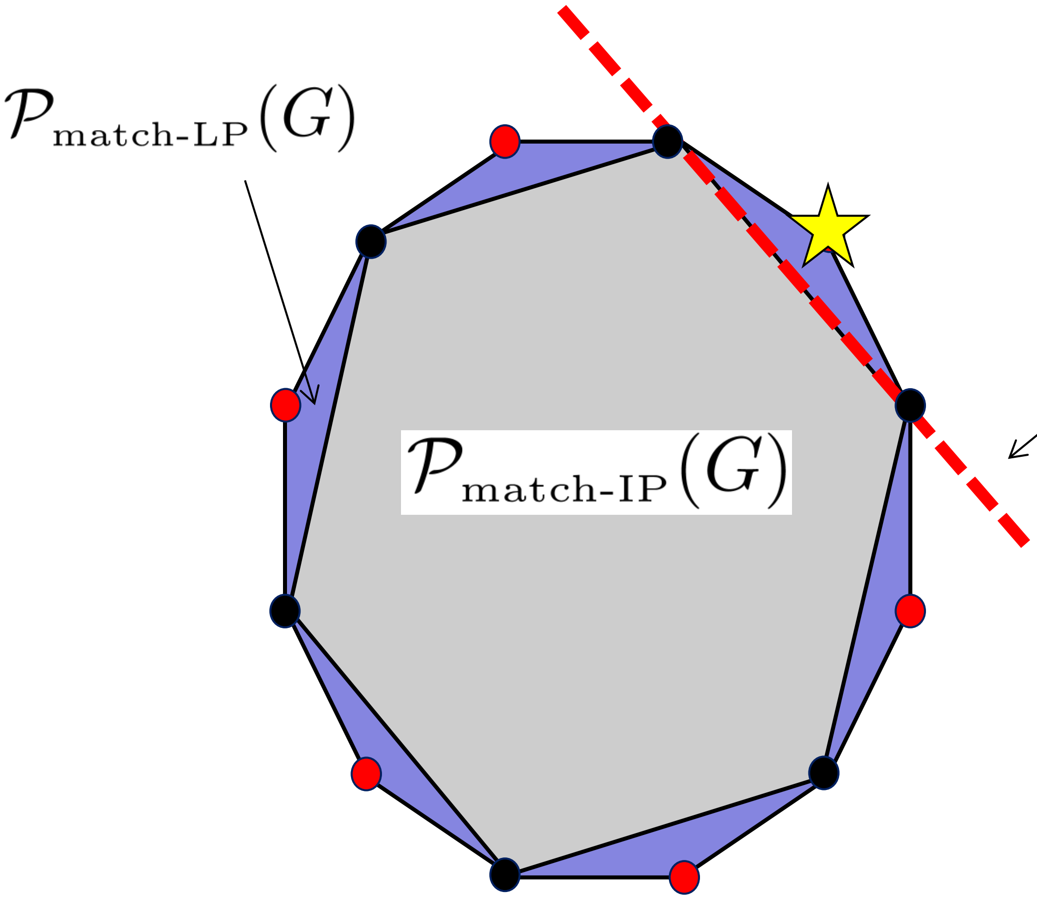
match-LP

What if match-LP is loose?



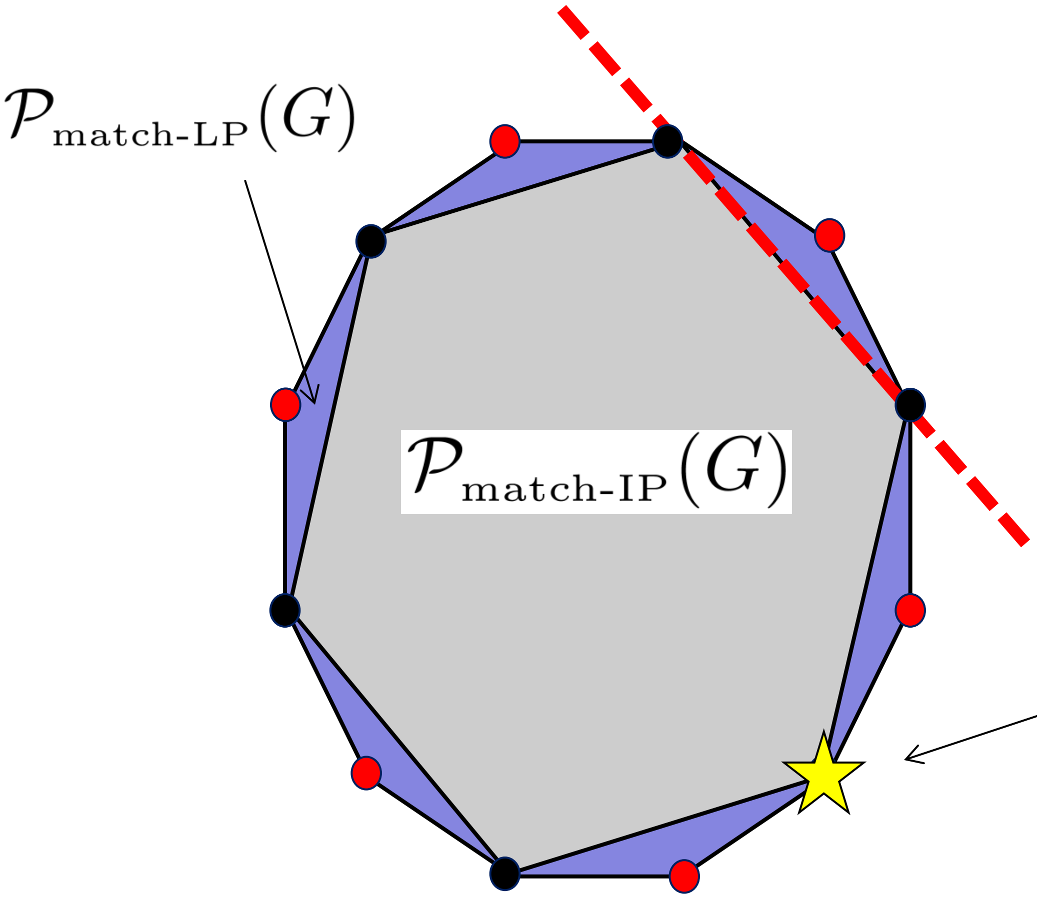
Solution of match-LP is fractional

What if match-LP is loose?



Remove fractional solution by adding constraints

What if match-LP is loose?



Solution of match-LP with additional cutting plane constraint

Edmonds and Blossom Constraints

- Matching polytope is also given by [Edmonds 65]:

$$\mathcal{P}_{\text{match-IP}}(G) = \mathcal{P}_{\text{match-blossom-LP}}(G)$$

$$\mathcal{P}_{\text{match-blossom-LP}}(G) = \Big\{\mathbf{x} \in [0,1]^{|E|} \mid \mathbf{x}(\delta(i)) \leq 1,\ \forall i \in V,$$

$$\sum_{e \in E(S)} x_e \leq \frac{|S|-1}{2},\ \forall S \in \mathcal{S}\Big\}$$

Edges w/ both ends in *S*

$$E(S) = \{(i,j) \in E \mid i,j \in S\}$$

All odd-sized subsets of *V*

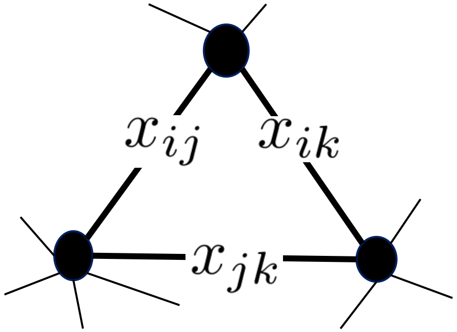
Edmonds and Blossom Constraints

- ☐ Matching polytope is also given by [Edmonds 65]:

$$\mathcal{P}_{\text{match-IP}}(G) = \mathcal{P}_{\text{match-blossom-LP}}(G)$$

$$\mathcal{P}_{\text{match-blossom-LP}}(G) = \Big\{ \mathbf{x} \in [0,1]^{|E|} \mid \mathbf{x}(\delta(i)) \le 1,\ \forall i \in V, \quad {\color{red} \sum x_e \le \frac{|S|-1}{2},\ \forall S \in \mathcal{S}} \Big\}$$

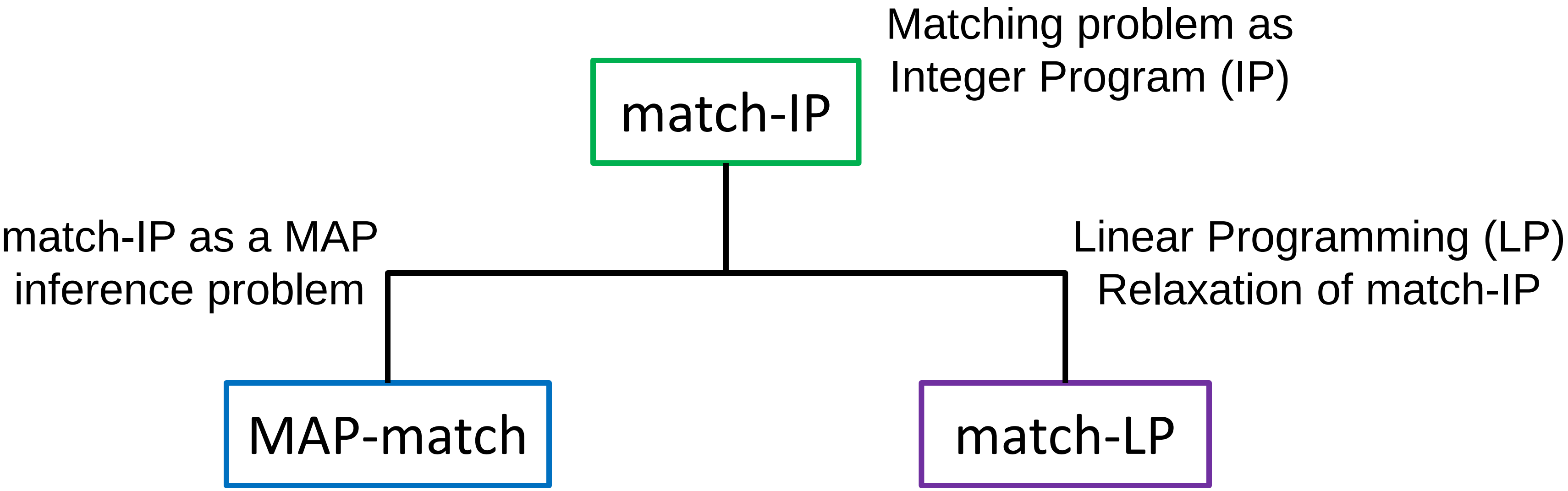
Subset of Size 3



Constraint: $x_{ij} + x_{ik} + x_{jk} \le 1$

Removes: $x_{ij} = x_{ik} = x_{jk} = \frac{1}{2}$

Solving Weighted Matching Problems



Matching problem as Integer Program (IP)

match-IP

match-IP as a MAP inference problem

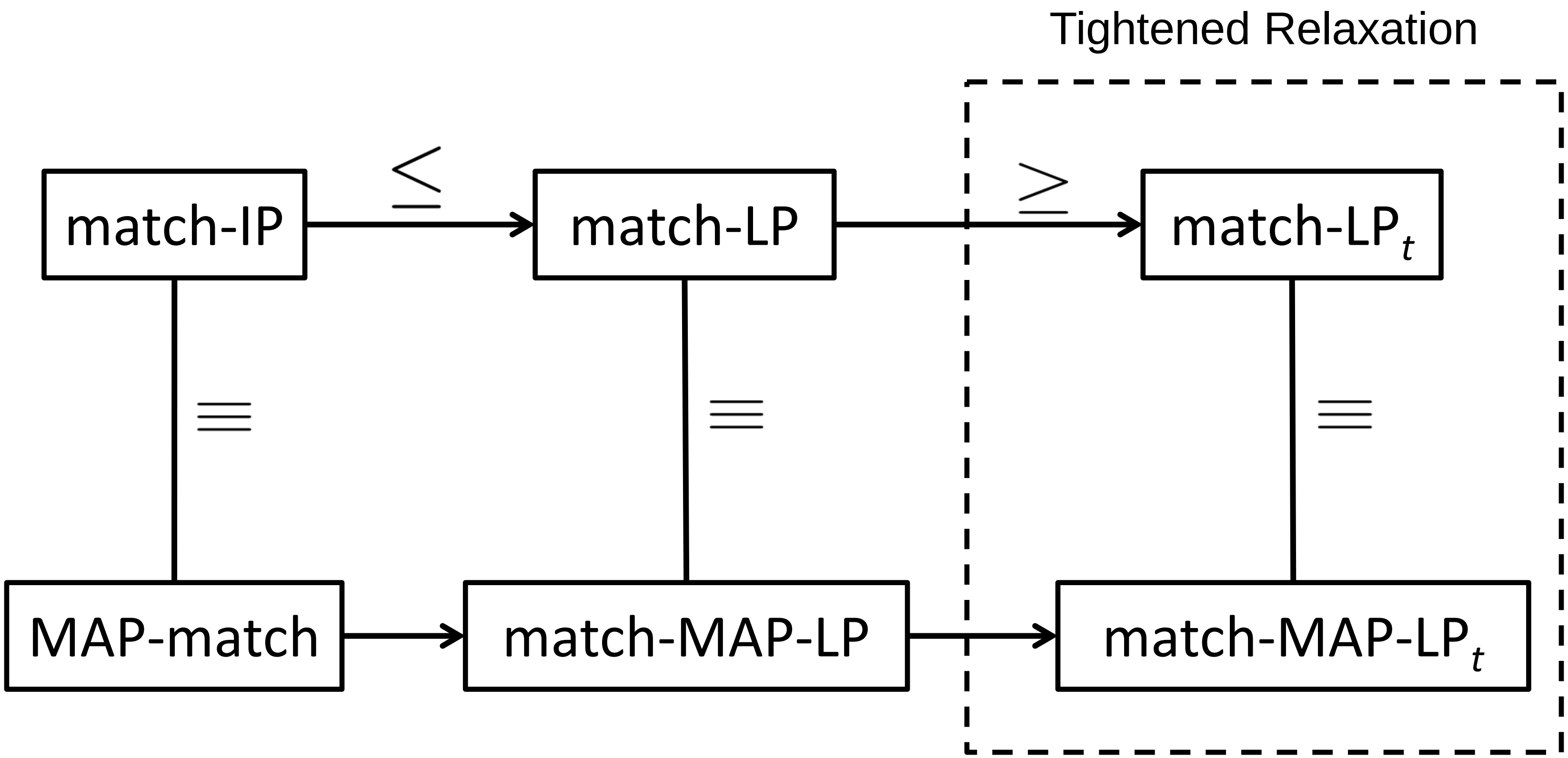
Linear Programming (LP) Relaxation of match-IP

MAP-match

match-LP

Bayati, Shah, Sharma [bipartite graphs]	2005	$O(c\,\lvert V\rvert^3)$
Sanghavi,Malioutov, Willsky [general graphs]	2007	$O(c\,w^{max}\lvert V\rvert^3)$

Overview of Results



Tightened Relaxation

match-IP $\xrightarrow{\leq}$ match-LP $\xrightarrow{\geq}$ match-LP$_t$

$\equiv$ $\equiv$ $\equiv$

MAP-match $\rightarrow$ match-MAP-LP $\rightarrow$ match-MAP-LP$_t$