ECE 6504: Advanced Topics in Machine Learning

Probabilistic Graphical Models and Large-Scale Learning

Topics

- Markov Random Fields: **MAP** Inference
  - Integer Programming, LP formulation
  - Dual Decomposition

Readings: KF 13.1-5, Barber 5.1,28.9

Dhruv Batra

Virginia Tech

Administrativia

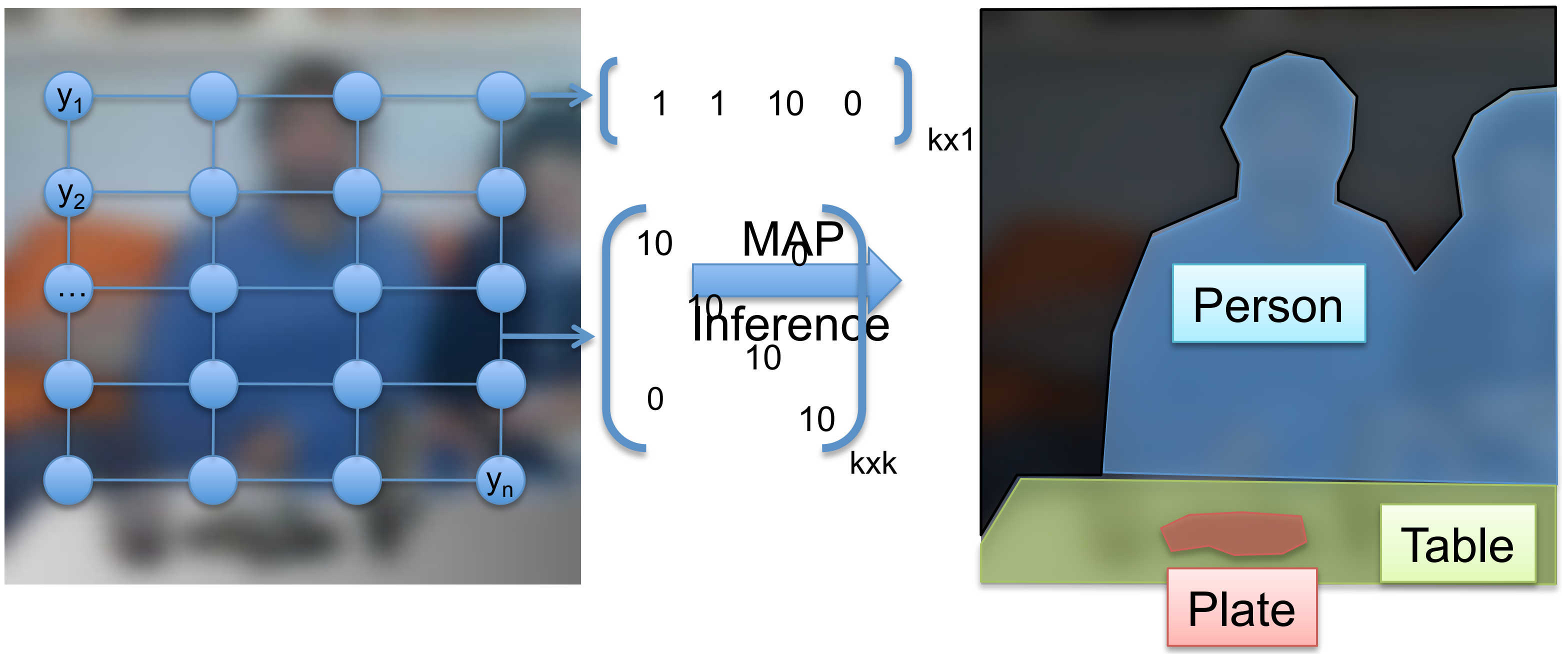
- HW1
  - Solutions and grades released

- HW2
  - Solutions released
  - Grades next week

- Project Presentations
  - When: April 22, 24
  - Where: in class
  - 5 min talk
    - Main results
    - Semester completion 2 weeks out from that point so nearly finished results expected
    - Slides due: April 21 11:55pm

Recap of Last Time

MAP Inference



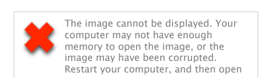
$$S(\mathbf{y}) = \sum_{i \in \mathcal{V}} \theta_i(y_i) + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}(y_i, y_j)$$

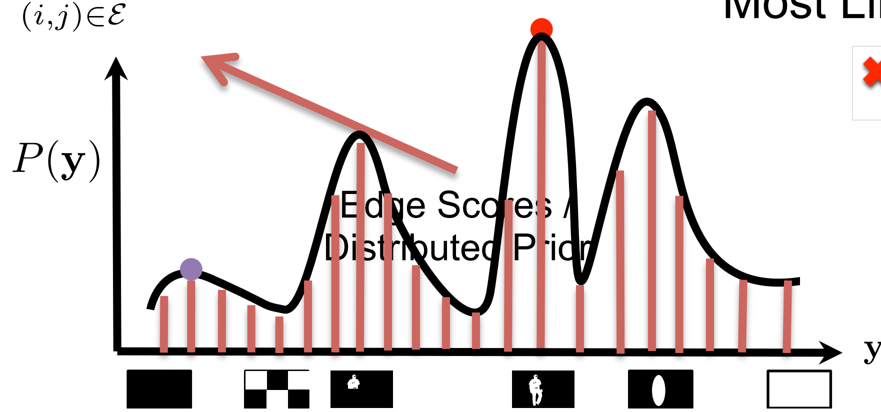
$$P(\mathbf{y}) = \frac{1}{Z} e^{S(\mathbf{y})}$$

Node Scores / Local Rewards

Edge Scores / Distributed Prior

Most Likely Assignment





MAP Inference

- Why is MAP difficult?

$$\text{MAP}(\theta) = \max_{\mathbf{x}} \sum_{i \in V} \theta_i(x_i) + \sum_{ij \in E} \theta_{ij}(x_i, x_j).$$

- What if we independently maximize the terms?

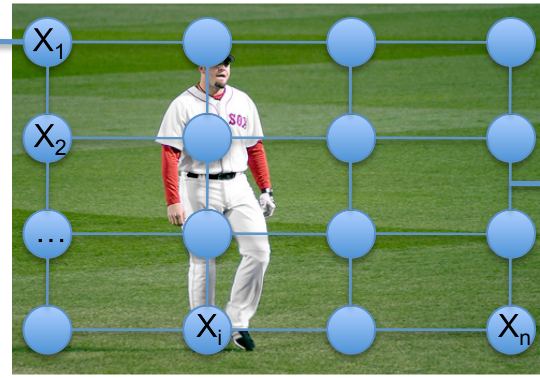
$$\text{MAP}(\theta) \leq \sum_{i \in V} \max_{x_i} \theta_i(x_i) + \sum_{ij \in E} \max_{x_i, x_j} \theta_{ij}(x_i, x_j)$$

MAP in Pairwise MRFs

$G = (\mathcal{V}, \mathcal{E})$

$$\begin{pmatrix} \theta_{ij}(1,1) & & \theta_{ij}(1,k) \\ & \cdots & \\ & & \cdots \\ \theta_{ij}(k,1) & & \theta_{ij}(k,k) \end{pmatrix}$$

$$\begin{pmatrix} \theta_1(1) \\ \cdots \\ \theta_1(k) \end{pmatrix}_{kx1}$$



$$\begin{pmatrix} \theta_{ij}(1,1) & & \theta_{ij}(1,k) \\ & \cdots & \\ & & \cdots \\ \theta_{ij}(k,1) & & \theta_{ij}(k,k) \end{pmatrix}_{kxk}$$

$$\begin{pmatrix} \theta_n(1) \\ \cdots \\ \theta_n(k) \end{pmatrix}$$

- Over-Complete Representation

x_1	$\cdots$	x_n	(x_1, x_2)	$\cdots$	(x_{n-1}, x_n)

$$\boldsymbol{\theta} = \Big[\theta_1(1)\ldots\theta_1(k) \quad \theta_n(1)\ldots\theta_n(k) \quad \theta_{12}(1,1)\ldots\theta_{12}(k,k) \quad \theta_{n-1,n}(1,1)\ldots\theta_{n-1,n}(k,k) \Big]$$

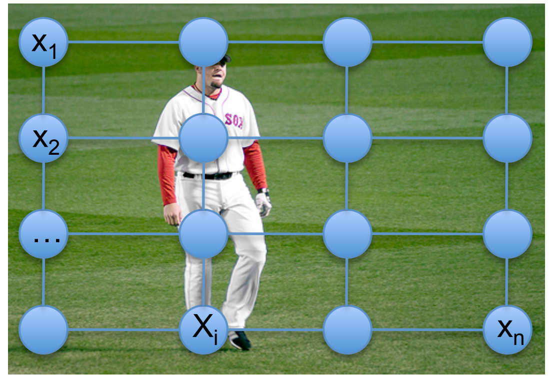
$\mu_1(1)\ldots\mu_1(k)$ $\qquad$ $\mu_n(1)\ldots\mu_n(k)$

$x_i \Rightarrow \begin{bmatrix} \cdot \\ \mu_i(s) \\ \cdot \\ \cdot \end{bmatrix}_{kx1}$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{x_i = 1} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}_{x_i = 2} \quad \times\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

MAP in Pairwise MRFs

$$G = (\mathcal{V}, \mathcal{E})$$

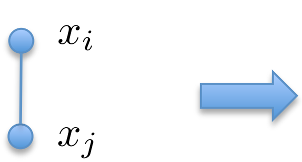


- Over-Complete Representation

$$\begin{array}{ccccccc} x_1 & \cdots & x_n & (x_1, x_2) & \cdots & (x_{n-1}, x_n) \end{array}$$

$$\boldsymbol{\theta} = \big[\theta_1(1)\ldots\theta_1(k) \quad \theta_n(1)\ldots\theta_n(k) \quad \theta_{12}(1,1)\ldots\theta_{12}(k,k) \quad \theta_{n-1,n}(1,1)\ldots\theta_{n-1,n}(k,k)\big]$$

$$\boldsymbol{\mu}_{\mathbf{x}} = \big[\mu_1(1)\ldots\mu_1(k) \quad \mu_n(1)\ldots\mu_n(k) \quad \mu_{12}(1,1)\ldots\mu_{12}(k,k) \quad \mu_{n-1,n}(1,1)\ldots\mu_{n-1,n}(k,k)\big]$$



$\mu_{ij}(s,t)$: k²x1

$$S(\mathbf{x}) = \boldsymbol{\theta} \cdot \boldsymbol{\mu}_{\mathbf{x}}$$

$[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^T$: $x_i = 1$, $x_j = 1$

$[0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^T$: $x_i = 1$, $x_j = 2$

MAP in Pairwise MRFs

- Integer Program

$$\max_{\boldsymbol{\mu}} \quad \boldsymbol{\theta}^T \boldsymbol{\mu}$$

$$\mu_i(s) \in \{0,1\}$$

$$\mu_{ij}(s,t) \in \{0,1\}$$

Indicator Variables

$$\sum_s \mu_i(s) = 1$$

$$\sum_{s,t} \mu_{i,j}(s,t) = 1$$

Unique Label

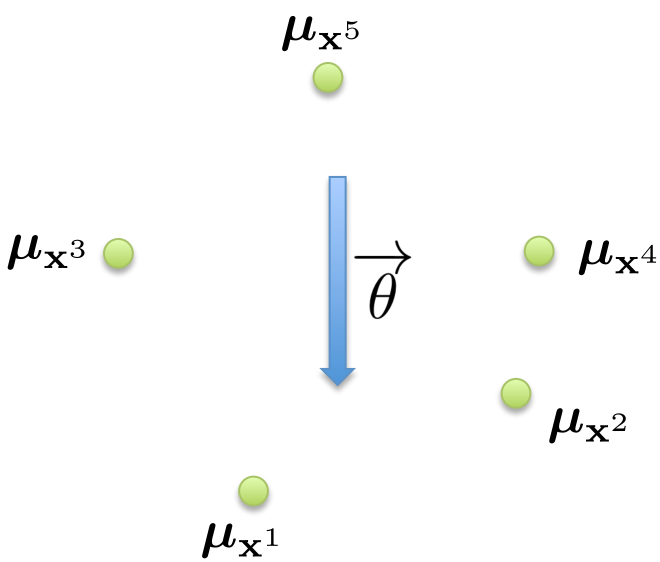
$$\sum_s \mu_{ij}(s,t) = \mu_j(t)$$

Consistent Assignments

MAP in Pairwise MRFs

- MAP Integer Program

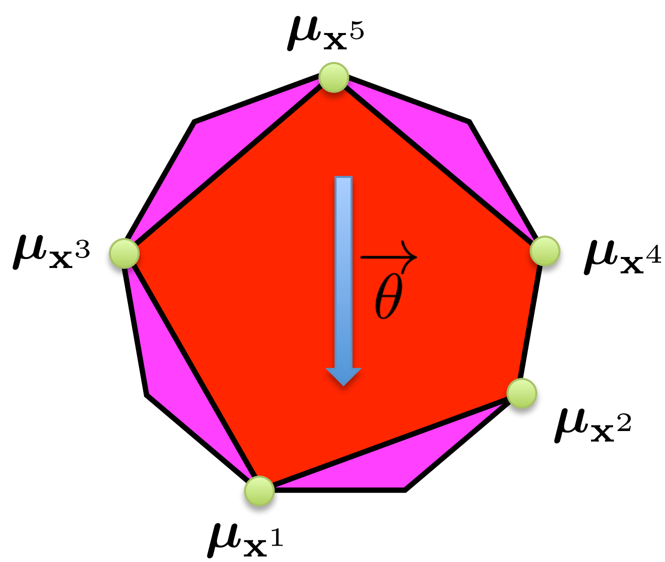
$$\max_{\boldsymbol{\mu}} \quad \boldsymbol{\theta}^T \boldsymbol{\mu}$$
$$s.t. \quad A\boldsymbol{\mu} = b$$
$$\boldsymbol{\mu}(\cdot) \in \{0, 1\}$$



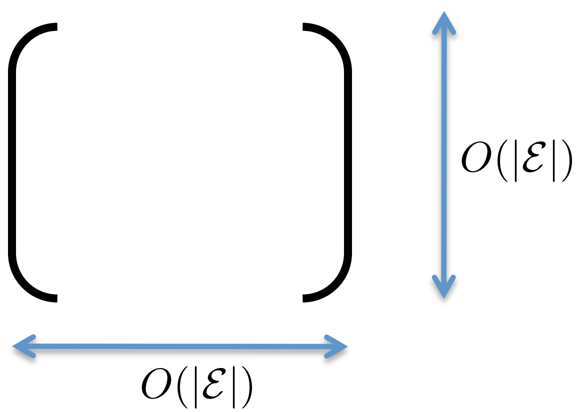
MAP in Pairwise MRFs

- MAP Linear Program

$$\max_{\boldsymbol{\mu}} \quad \boldsymbol{\theta}^T \boldsymbol{\mu}$$
$$s.t. \quad A\boldsymbol{\mu} = b$$
$$\boldsymbol{\mu}(\cdot) \in [0, 1]$$



$A =$



Off-the-shelf solvers
CPLEX
Mosek
etc

✗

Plan for today

- MRF Inference
  - (Specialized) MAP Inference
    - Integer Programming Formulation
    - Linear Programming Relaxation
      - Understanding the LP better
      - When is it tight?
      - When is it not?
    - Dual Decomposition
      - Algorithm for solving this LP

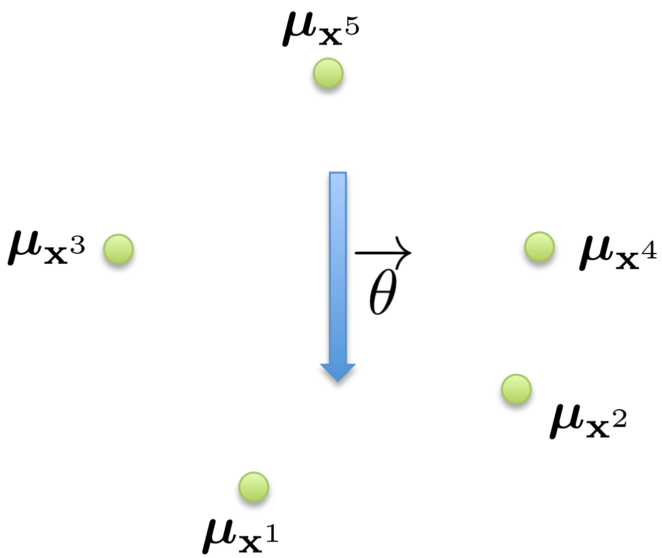
MAP in Pairwise MRFs

- MAP Integer Program

$$\max_{\boldsymbol{\mu}} \quad \boldsymbol{\theta}^T \boldsymbol{\mu}$$

$$s.t. \quad A\boldsymbol{\mu} = b$$

$$\boldsymbol{\mu}(\cdot) \in \{0, 1\}$$



Marginal Polytope

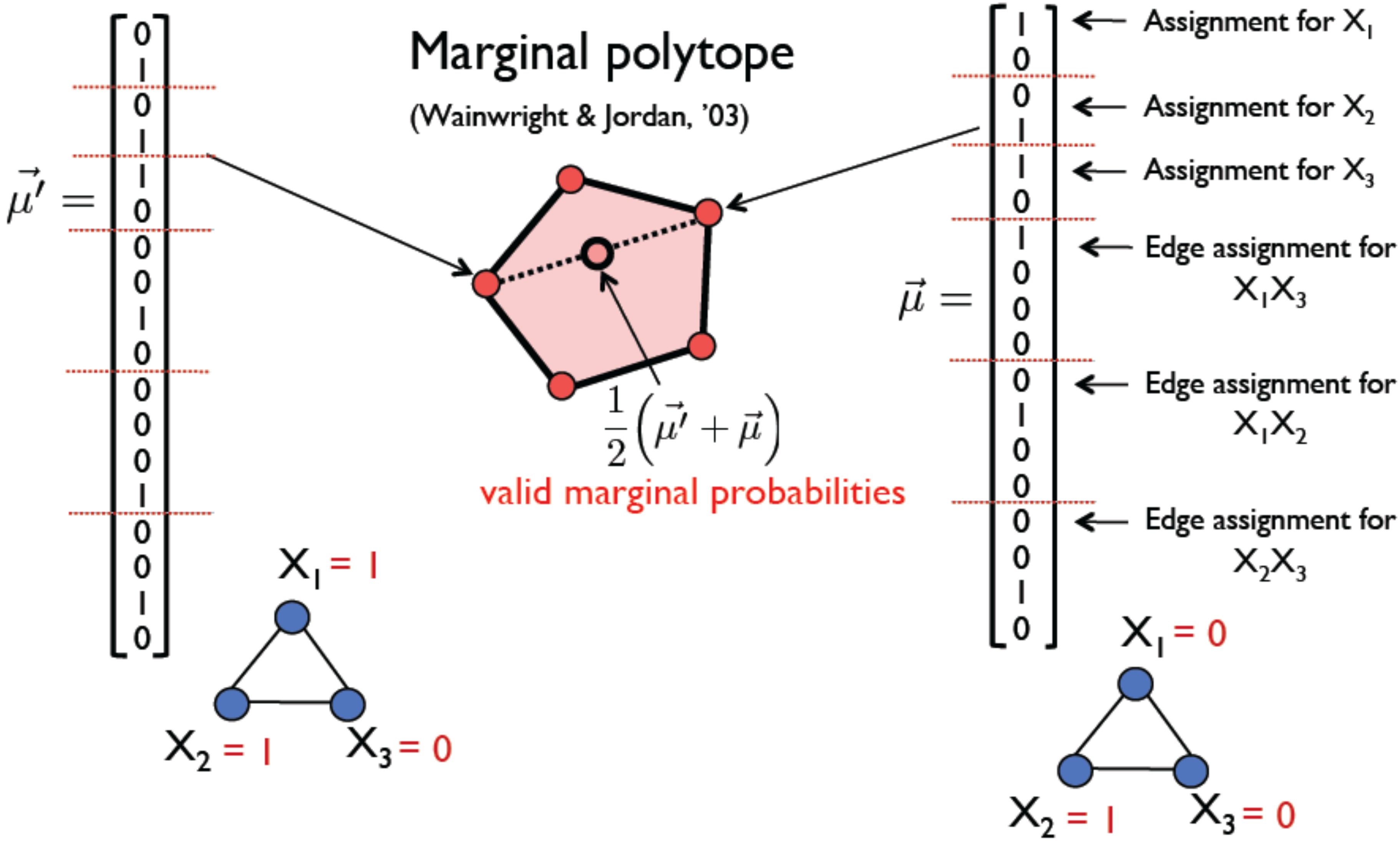
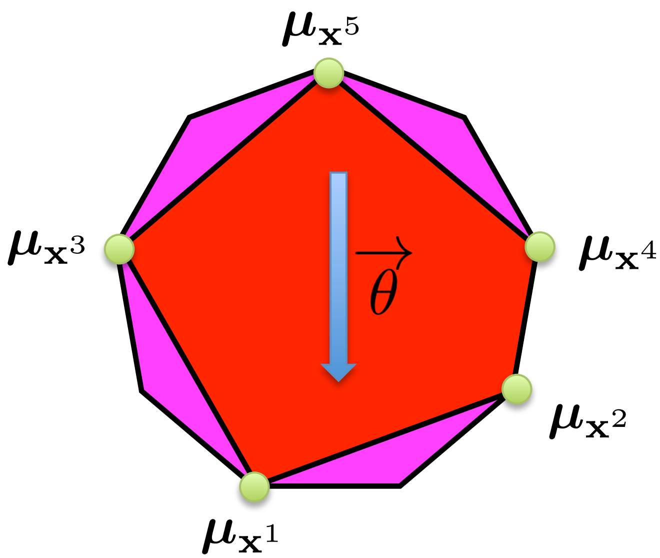


Figure Credit: David Sontag

MAP in Pairwise MRFs

- MAP Linear Program

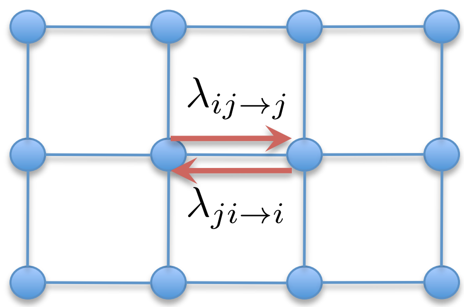
$$\max_{\boldsymbol{\mu}} \quad \boldsymbol{\theta}^T \boldsymbol{\mu}$$
$$s.t. \quad A\boldsymbol{\mu} = b$$
$$\boldsymbol{\mu}(\cdot) \in [0, 1]$$



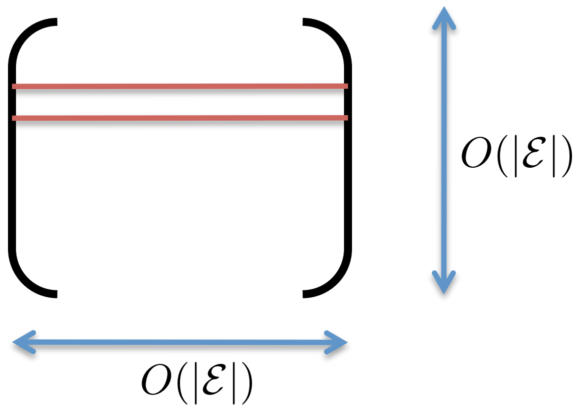
- Properties
  - If LP-opt is integral, MAP is found
  - LP always integral for trees
  - Efficient message-passing schemes for solving LP

LP Relaxation

- Block Co-ordinate / Sub-gradient Descent on Dual

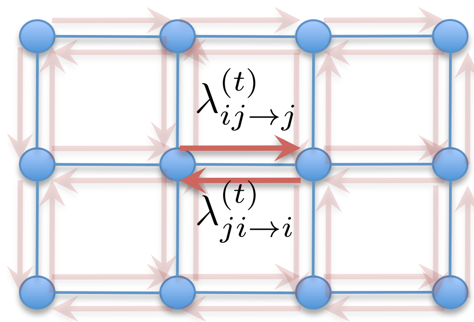


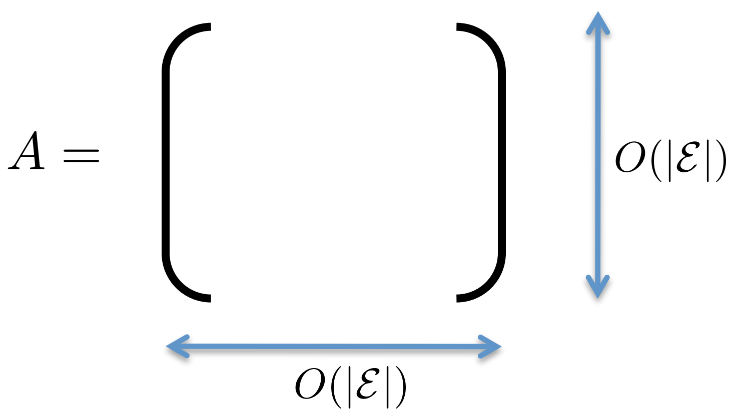
$A =$



LP Relaxation

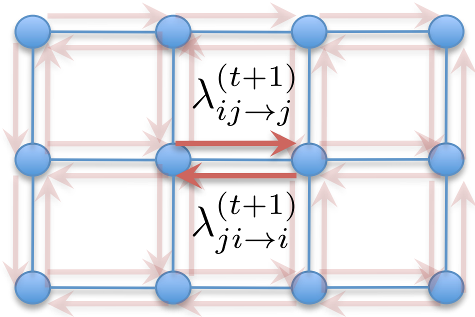
- Block Co-ordinate / Sub-gradient Descent on Dual





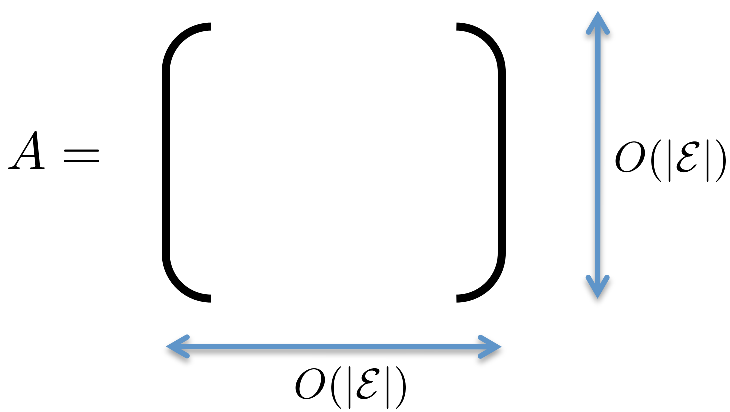
LP Relaxation

- Block Co-ordinate / Sub-gradient Descent on Dual

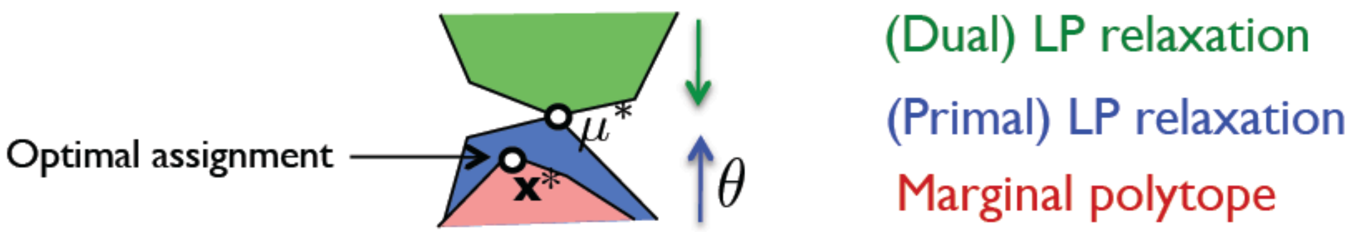


Distributed Message-Passing

Still inefficient!



Linear Programming Duality



Dual Decomposition

- For MAP Inference
  - On board

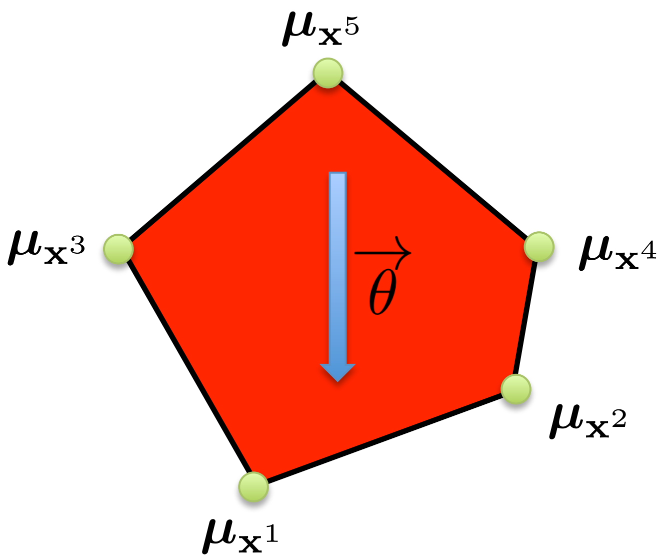
MAP in Pairwise MRFs

- MAP Integer Program

$$\max_{\boldsymbol{\mu}} \quad \boldsymbol{\theta}^T \boldsymbol{\mu}$$

$$s.t. \quad A\boldsymbol{\mu} = b$$

$$\boldsymbol{\mu}(\cdot) \in \{0, 1\}$$



MAP LP

- Lagrangian Relaxation

$$f(\lambda) = \max_{\boldsymbol{\mu} \in \mathcal{C}} \sum_i \boldsymbol{\theta}_i \cdot \boldsymbol{\mu}_i + \sum_{(i,j)} \boldsymbol{\theta}_{ij} \cdot \boldsymbol{\mu}_{ij} \quad -\lambda \cdot (A\boldsymbol{\mu} - b)$$

$$s.t. \quad \boldsymbol{\mu}_i(\cdot), \boldsymbol{\mu}_{ij}(\cdot) \in \{0,1\}$$

Dual $\min_{\lambda \geq 0} \quad f(\lambda)$