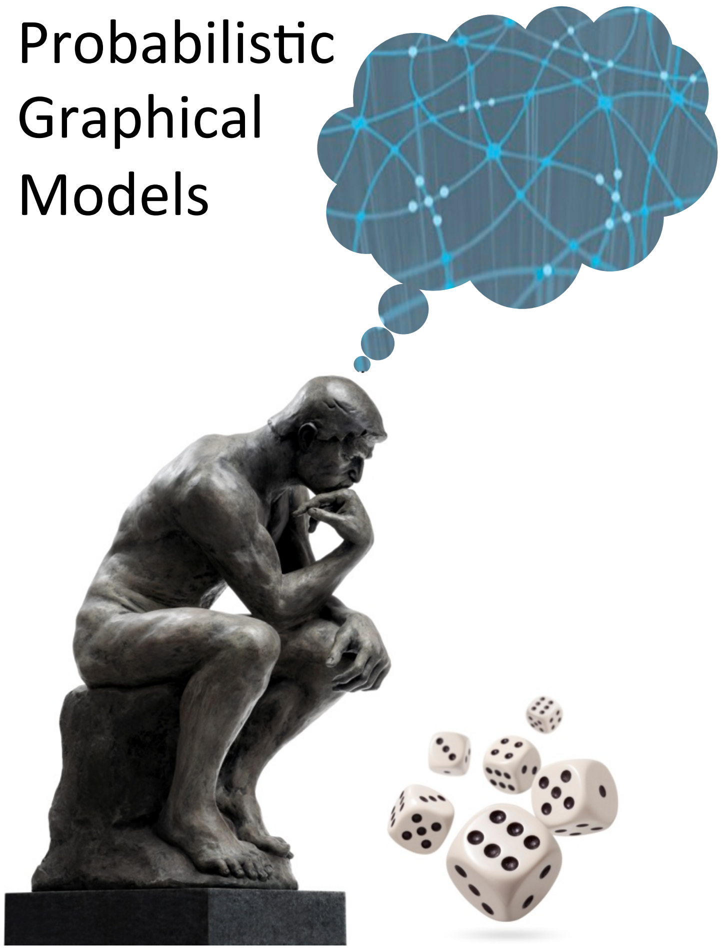
Probabilistic
Graphical
Models



Inference

MAP

# Dual Decomposition Algorithm

# Dual Decomposition Algorithm

$$\bar{\theta}_i^{\boldsymbol{\lambda}} = \theta_i(x_i) + \sum_{F:i\in F} \lambda_{Fi}(x_i) \qquad \bar{\theta}_F^{\boldsymbol{\lambda}} = \theta_F(\boldsymbol{x}_F) - \sum_{i\in F} \lambda_{Fi}(x_i)$$

$x_i$ F

- Initialize all λ's to be 0
- Repeat for t=1,2,...
  - Locally optimize all slaves: $\boldsymbol{x}_F^* = \text{argmax}_{\boldsymbol{x}_F} \bar{\theta}_F^{\boldsymbol{\lambda}}(\boldsymbol{x}_F)$
  $x_i^* = \text{argmax}_{x_i} \bar{\theta}_i^{\boldsymbol{\lambda}}(x_i)$
  - For all F and i∈F

disagree
    - If $x_{Fi}^* \neq x_i^*$ then

$\alpha_t > 0$

$$\lambda_{Fi}(x_i^*) := \lambda_{Fi}(x_i^*) - \alpha_t$$
$$\lambda_{Fi}(x_{Fi}^*) := \lambda_{Fi}(x_{Fi}^*) + \alpha_t$$

Daphne Koller

# Dual Decomposition Convergence

- Under weak conditions on $\alpha_t$, the $\lambda$'s are guaranteed to converge
  - $\Sigma_t \alpha_t = \infty$
  - $\Sigma_t \alpha_t^2 < \infty$
- Convergence is to a unique global optimum, regardless of initialization

# At Convergence

- Each slave has a locally optimal solution over its own variables (in its scope)
- Solutions may not agree on shared variables
- If all slaves agree, the shared solution is a guaranteed MAP assignment
- Otherwise, we need to solve the decoding problem to construct a joint assignment

# Options for Decoding x*

- Several heuristics
  - If we use decomposition into spanning trees, can take MAP solution of any tree
  - Have each slave vote on $X_i$'s in its scope & for each $X_i$ pick value with most votes
  - Weighted average of sequence of messages sent regarding each $X_i$
- Score $\theta$ is easy to evaluate for any x
- **Best to generate many candidates and pick the one with highest score**

Daphne Koller

# Upper Bound

- $L(\lambda)$ is upper bound on $MAP(\theta)$

$$\underbrace{\overset{\text{candidate MAP}}{\text{score}(\mathbf{x})}}_{\theta(x)} \leq \overset{?}{MAP(\theta)} \leq \underline{L(\lambda)}$$

$$\underbrace{MAP(\theta) - \text{score}(\mathbf{x})} \leq \underbrace{L(\lambda) - \text{score}(\mathbf{x})}_{\text{small enough}}$$

# Important Design Choices

- Division of problem into slaves
  - Larger slaves (with more factors) improve convergence and often quality of answers
- Selecting locally optimal solutions for slaves
  - Try to move toward faster agreement
- Adjusting the step size $\alpha_t$
- Methods to construct candidate solutions

# Summary: Algorithm

- Dual decomposition is a general-purpose algorithm for MAP inference
  - Divides model into tractable components
  - Solves each one locally
  - Passes "messages" to induce them to agree
- Any tractable MAP subclass can be used in this setting as a slave

# Summary: Theory

- Formally: a subgradient optimization algorithm on dual problem to MAP
- Provides important guarantees
  - Upper bound on distance to MAP
  - Conditions that guarantee exact MAP solution
- Even some analysis for which decomposition into slaves is better

# Summary: Practice

- Pros:
  - Very general purpose
  - Best theoretical guarantees
  - Can use very fast, specialized MAP subroutines for solving large model components
- Cons:
  - Not the fastest algorithm
  - Lots of tunable parameters / design choices