Paired-Dual Learning for Fast Training of Latent Variable Hinge-Loss MRFs

Stephen H. Bach* Maryland
Bert Huang* Virginia Tech
Jordan Boyd-Graber Colorado
Lise Getoor UC Santa Cruz

* Equal Contributors

Introduction

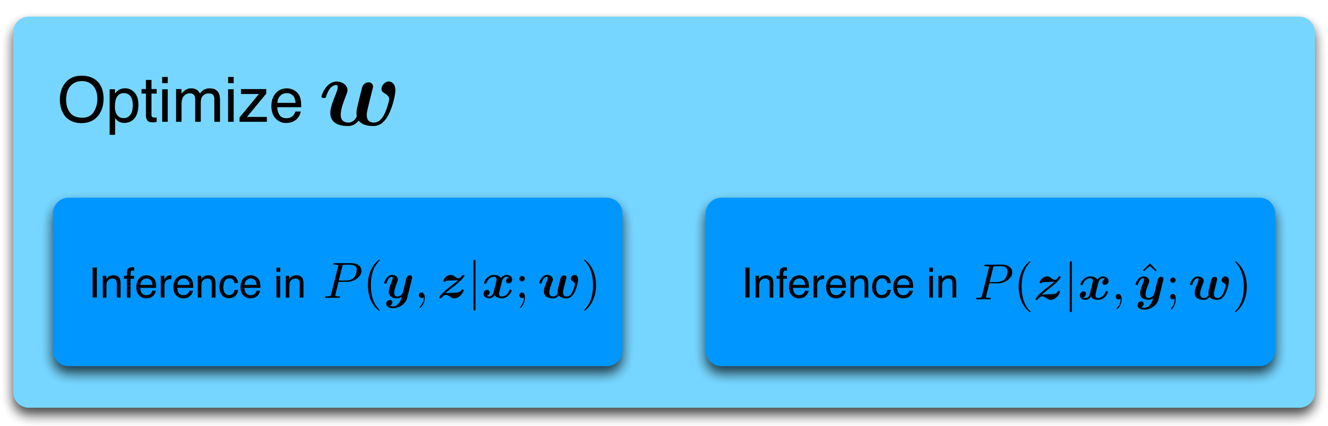
- **Latent variables** are very useful in **large-scale structured prediction**, but they **complicate learning**!
- A naive approach is not scalable, and smarter methods exist **for discrete models only**
- We introduce a **new method for continuous MRFs** that interleaves parameter and inference updates

Learning with Latent Variables

- The maximum likelihood learning objective for models with latent variables contains two inner inference problems:

$$\arg\min_{\boldsymbol{w}} \max_{\rho \in \Delta(\boldsymbol{y},\boldsymbol{z})} \min_{q \in \Delta(\boldsymbol{z})} \frac{\lambda}{2}\|\boldsymbol{w}\|^2 - \mathbb{E}_\rho\left[\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\right] + H(\rho) + \mathbb{E}_q\left[\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x},\hat{\boldsymbol{y}},\boldsymbol{z})\right] - H(q)$$

- At a high level, this objective has a simple structure:

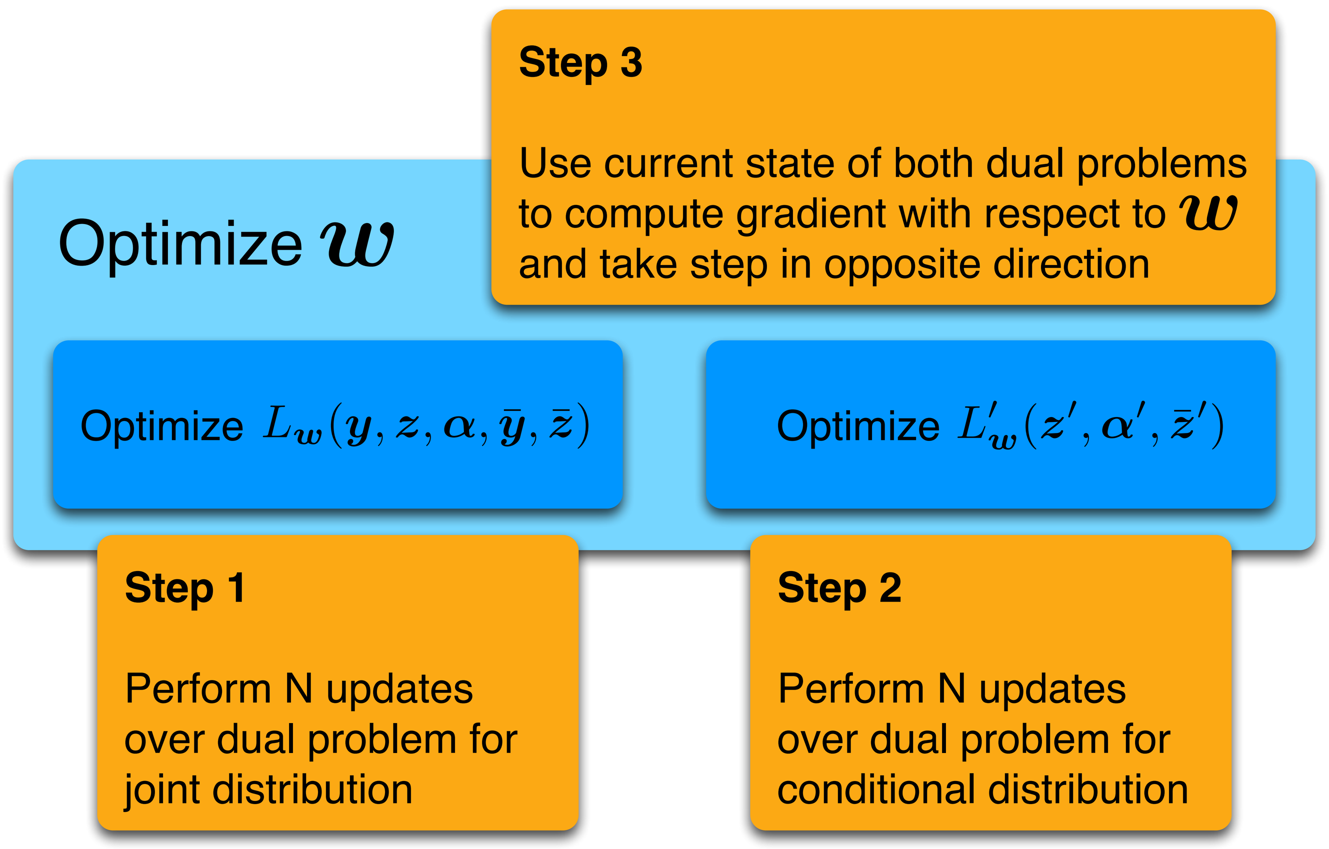


- But repeatedly performing inference is **very expensive**!
- Supervised learning can be sped up by interleaving inference and parameter updates, e.g., Taskar et al. [ICML 2005] and Meshi et al. [ICML 2010]
- Schwing et al. [ICML 2012] and Chen et al. [ICML 2015] propose interleaving updates for discrete latent models
- Any new method for continuous variables must solve the problems of **intractable expectations and entropies**

Paired-Dual Learning

- We choose point distributions for variational families and construct entropy surrogates to make objective tractable
- Paired-dual learning replaces both inner inferences with augmented Lagrangians and optimizes jointly:

$$\arg\min_{\boldsymbol{w}} \max_{\boldsymbol{y},\boldsymbol{z},\bar{\boldsymbol{y}},\bar{\boldsymbol{z}}} \min_{\boldsymbol{\alpha}} \min_{\boldsymbol{z}',\bar{\boldsymbol{z}}'} \max_{\boldsymbol{\alpha}'} \frac{\lambda}{2}\|\boldsymbol{w}\|^2 - L_{\boldsymbol{w}}(\boldsymbol{y},\boldsymbol{z},\boldsymbol{\alpha},\bar{\boldsymbol{y}},\bar{\boldsymbol{z}}) + L'_{\boldsymbol{w}}(\boldsymbol{z}',\boldsymbol{\alpha}',\bar{\boldsymbol{z}}')$$



Hinge-Loss Markov Random Fields

- Undirected graphical models over continuous variables with hinge-loss potential functions

$$P(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{x};\boldsymbol{w}) \propto \exp\left(\sum_{j=1}^{m} w_j \left(\max\{\ell_j(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}),0\}\right)^{p_j}\right)$$

where ℓ_j is a linear function and $p_j \in \{1,2\}$

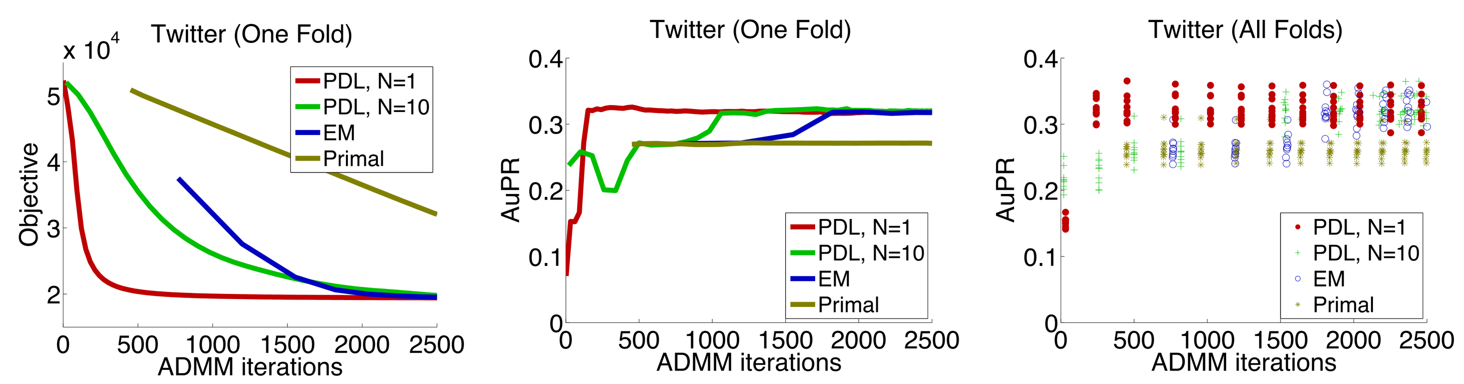
- Generalizes
  - Randomized algorithms for MAX SAT
  - Relaxed MAP for discrete, logical MRFs
  - Exact MAX SAT using soft logic
- **Highly scalable**, ADMM-based MAP inference

Code available: http://psl.cs.umd.edu

Evaluation

- Paired-dual learning is
  - **Just as accurate** as traditional methods
  - **So fast** that it often **converges before others make a single parameter update**

Interaction Prediction via Community Detection



Trust Prediction in Social Networks

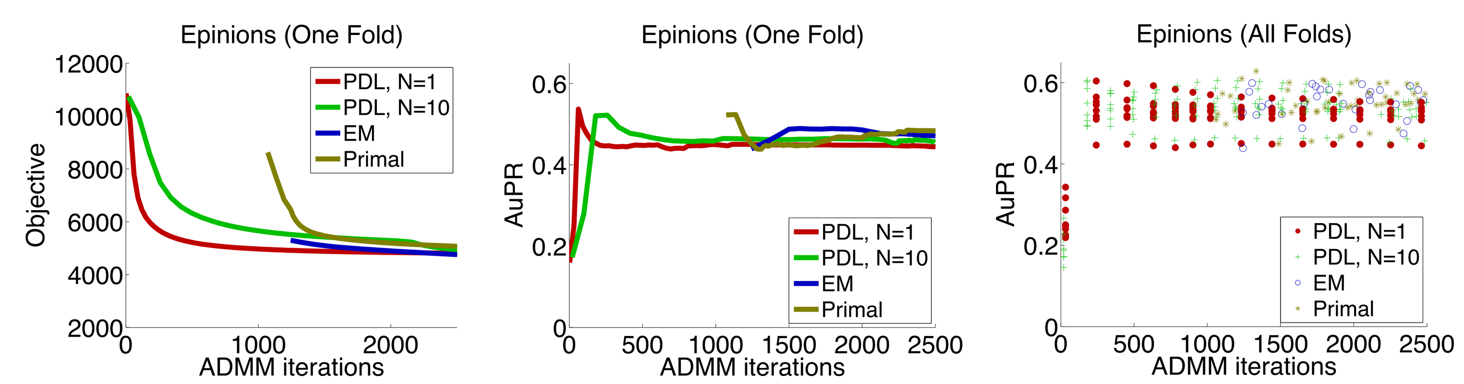
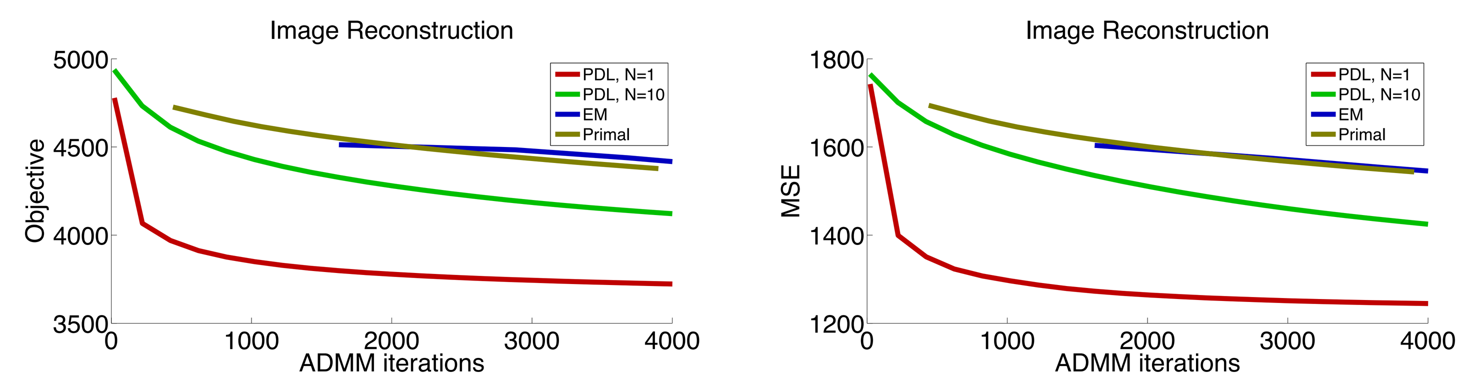


Image Reconstruction



Conclusions

- Paired-dual learning overcomes the inference bottleneck associated with learning with latent variables
- Latent variable hinge-loss MRFs are now **practical for large-scale applications!**
- Paired-dual learning is also applicable to other continuous models and even discrete models

This work is supported by NSF grant IIS1218488 and IARPA via DoI/NBC contract number D12PC00337. Boyd-Graber is supported by NSF grants IIS1320538, IIS1409287, and NCSE1422492. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.