A Combined LP and QP Relaxation for MAP

Patrick Pletscher, Sharon Wulff

Machine Learning Laboratory, ETH Zürich, Switzerland

NIPS DISCML 2011 Workshop, Sierra Nevada

MAP Inference for General Pairwise Energies

Energy Minimization

Discrete pairwise energy minimization for graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

$$\min_{\mathbf{x}} \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}(x_i, x_j) \qquad x_i \in \{1, \ldots, K\}.$$

For general $\mathcal{G}$ and θ this is an NP-hard problem.

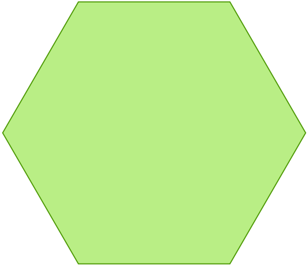
Linear and Quadratic Programming Relaxations

Ravikumar and Lafferty, 2006; Wainwright and Jordan, 2008

LP for Marginal Polytope

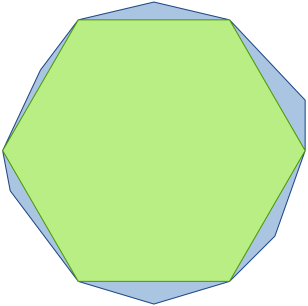
$$\min_{\boldsymbol{\mu} \in \mathcal{M}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^\mathsf{T} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^\mathsf{T} \boldsymbol{\mu}_{ij}$$

But: $\mathcal{M}_{\mathcal{G}}$ is exponentially large



Linear and Quadratic Programming Relaxations

Ravikumar and Lafferty, 2006; Wainwright and Jordan, 2008



LP for Marginal Polytope

$$\min_{\boldsymbol{\mu} \in \mathcal{M}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}$$

But: $\mathcal{M}_{\mathcal{G}}$ is exponentially large

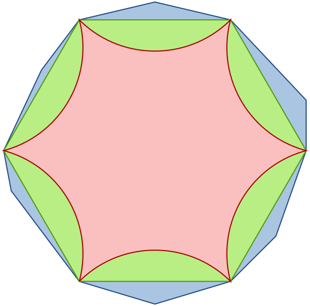
outer approximation

LP for Local Marginal Polytope

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}$$

Linear and Quadratic Programming Relaxations

Ravikumar and Lafferty, 2006; Wainwright and Jordan, 2008



LP for Marginal Polytope

$$\min_{\boldsymbol{\mu} \in \mathcal{M}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}$$

But: $\mathcal{M}_{\mathcal{G}}$ is exponentially large

outer approximation

inner approximation

LP for Local Marginal Polytope

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}$$

Quadratic Programming

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}$$

$$\text{s.t.} \quad \boldsymbol{\mu}_{ij} = \text{vec}(\boldsymbol{\mu}_i \boldsymbol{\mu}_j^{\mathsf{T}}) \quad \forall (i,j) \in \mathcal{E}$$

Our Approach in a Nutshell

- Combine Linear and Quadratic Programming relaxations.
- Kullback-Leibler penalty term to discourage inconsistencies.
- Solved using an efficient belief-propagation algorithm.

Combined Formulation

Joint Objective

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \underbrace{\sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}}_{\text{standard LP objective}} + \underbrace{\beta \sum_{(i,j) \in \mathcal{E}} D_{KL}(\boldsymbol{\mu}_{ij}, \text{vec}(\boldsymbol{\mu}_i \boldsymbol{\mu}_j^{\mathsf{T}}))}_{\text{penalty term for inconsistencies}} .$$

- Constraint set: local marginal polytope $\mathcal{L}_{\mathcal{G}}$.
- Kullback-Leibler divergence as the penalty term:

$$D_{KL}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_k \mu_k \log \left(\frac{\mu_k}{\nu_k} \right) .$$

- Limit cases for β
  - $\beta = 0$: LP relaxation
  - $\beta \to \infty$: QP relaxation.
- Non-convex formulation.

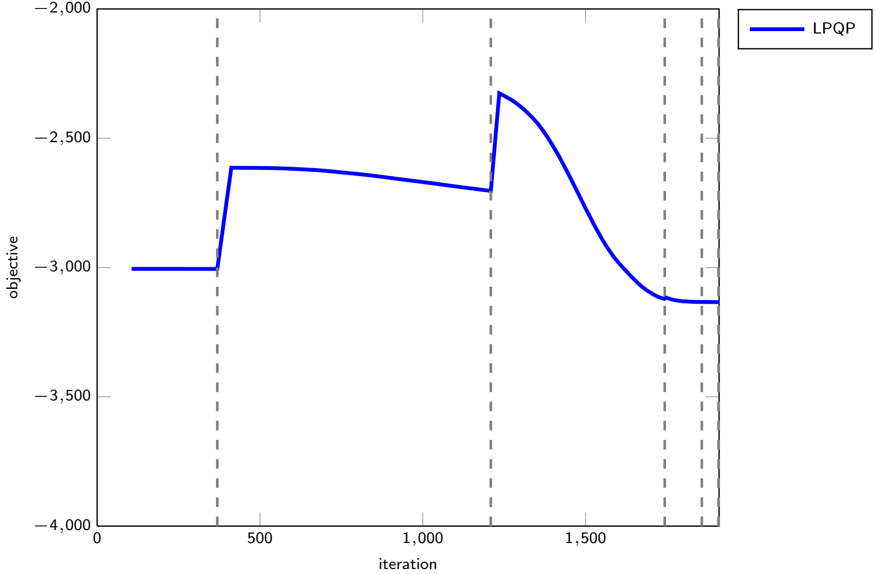
LPQP Algorithm Highlights

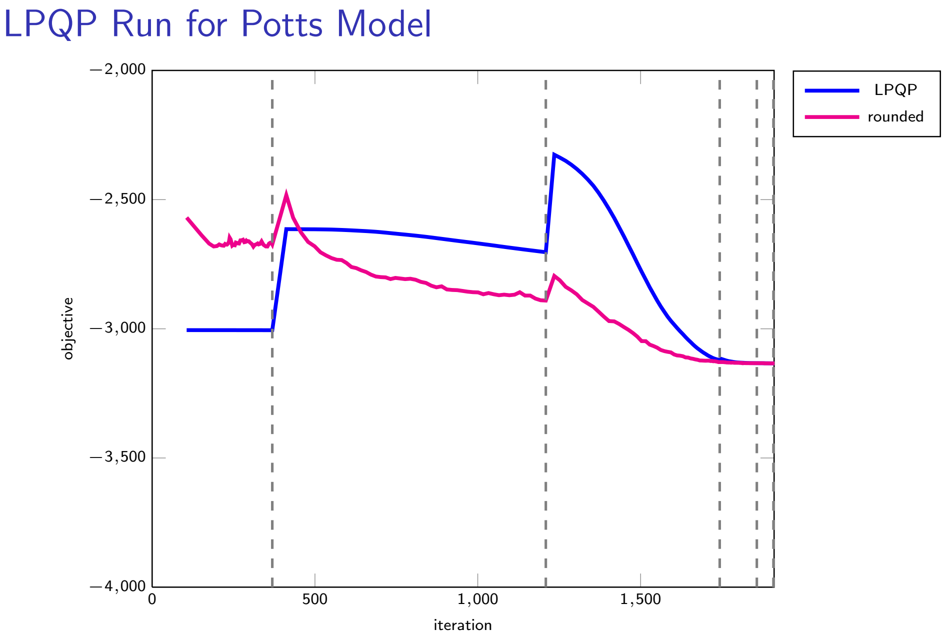
- Decompose objective into a difference of convex functions.
- The main computational task is the following convex problem

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \underbrace{\sum_{i \in \mathcal{V}} \tilde{\boldsymbol{\theta}}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}}_{\text{modified LP relaxation}} - \beta \underbrace{\sum_{(i,j) \in \mathcal{E}} H(\boldsymbol{\mu}_{ij})}_{\text{pairwise entropy}} .$$

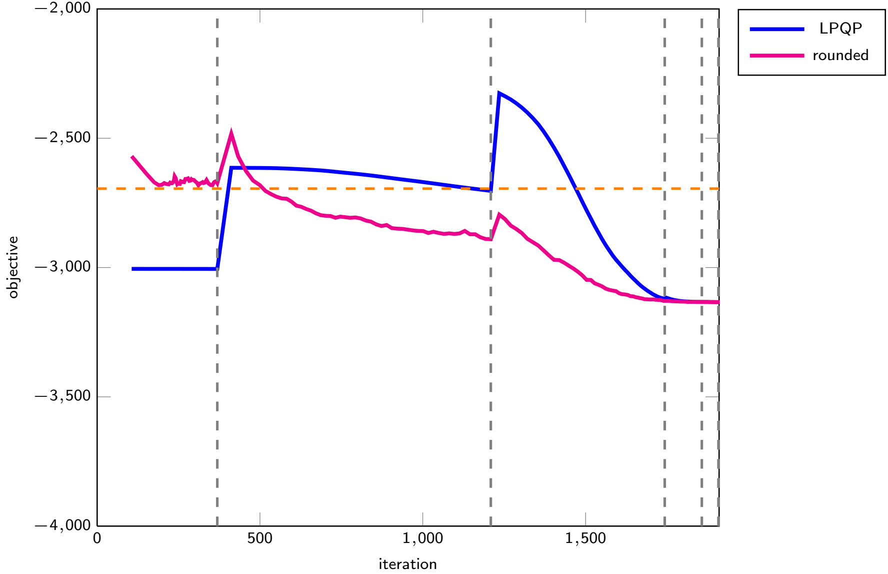
- Solved by norm-product algorithm (Hazan and Shashua, 2010).
- Can warm-start with solution from previous iteration.
- Annealing for β.

LPQP Run for Potts Model

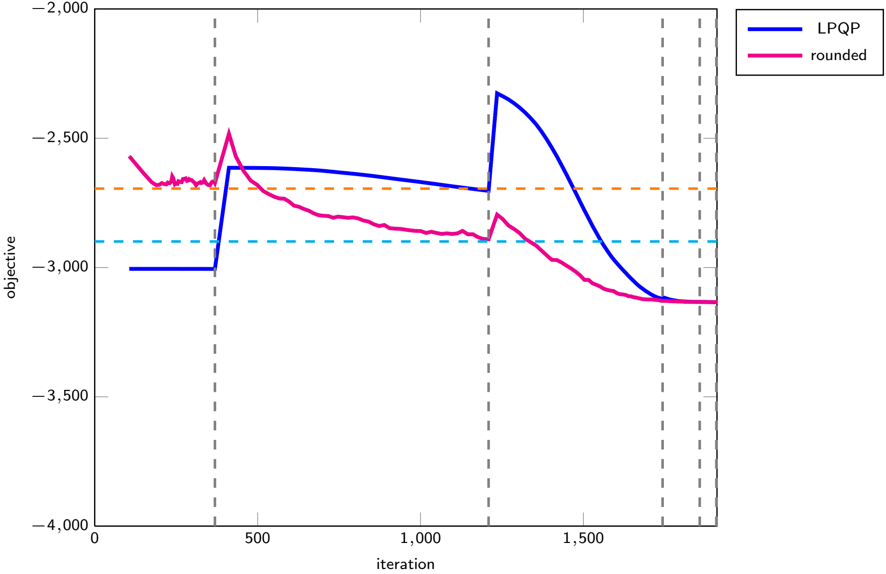



LPQP Run for Potts Model
−2,000
−2,500
−3,000
−3,500
−4,000
0
500
1,000
1,500
objective
iteration
LPQP
rounded

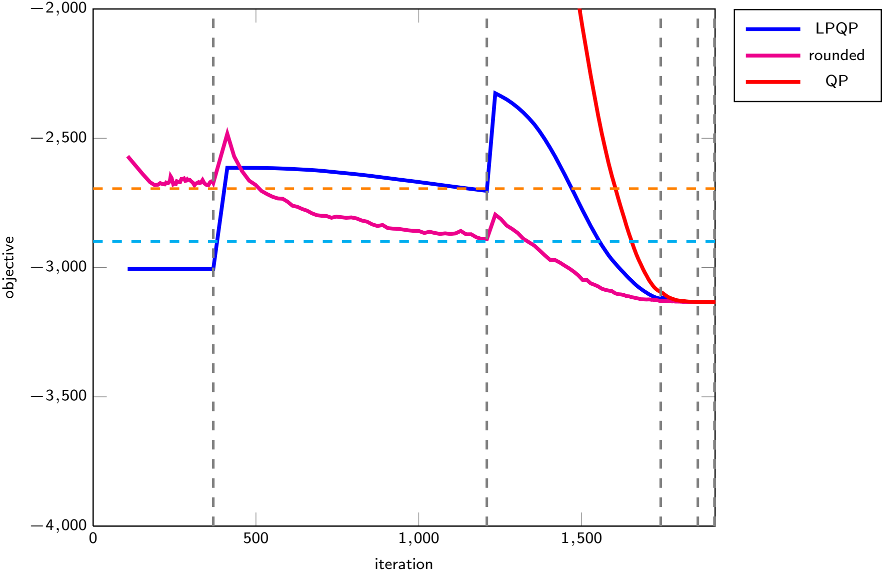
LPQP Run for Potts Model



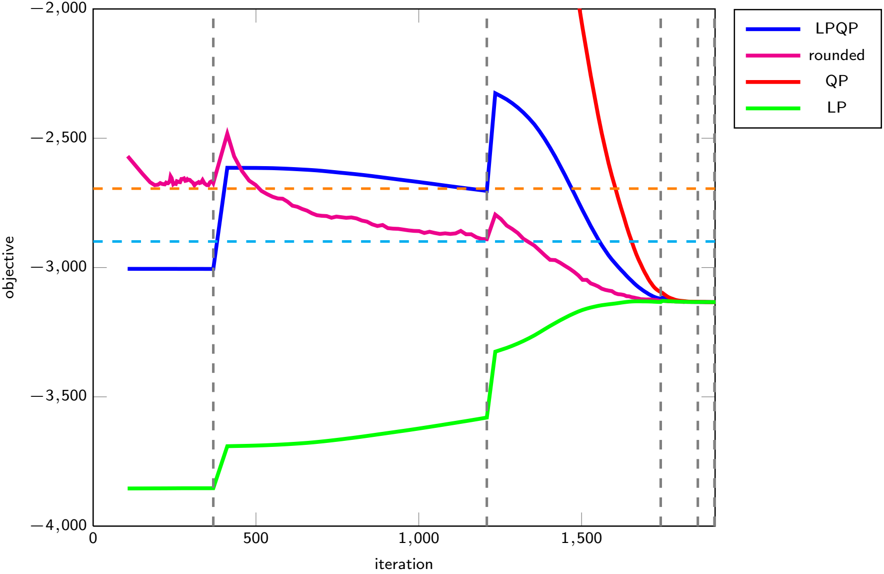
LPQP Run for Potts Model



LPQP Run for Potts Model



LPQP Run for Potts Model



References I

Hazan, Tamir and Amnon Shashua (2010). "Norm-Product Belief Propagation: Primal-Dual Message-Passing for Approximate Inference". In: *IEEE Transactions on Information Theory* 56.12, pp. 6294–6316.

Ravikumar, Pradeep and John Lafferty (2006). "Quadratic Programming Relaxations for Metric Labeling and Markov Random Field MAP Estimation". In: *ICML*.

Wainwright, Martin J and Michael I Jordan (2008). *Graphical Models, Exponential Families, and Variational Inference*. Vol. 1.