LPQP for MAP: Putting LP Solvers to Better Use

Patrick Pletscher, Sharon Wulff

Machine Learning Laboratory, ETH Zürich, Switzerland

ICML
June 28, 2012

MAP Inference for General Pairwise Energies

Energy Minimization

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

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

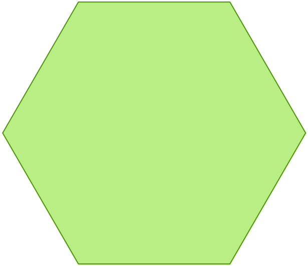
For general $\mathcal{G}$ and θ this is an NP-hard problem.

Linear and Quadratic Programming Relaxations

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

LP for Marginal Polytope

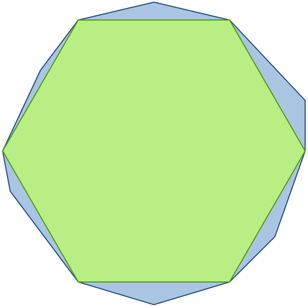
$$\min_{\boldsymbol{\mu} \in \mathcal{M}_{\mathcal{G}}} \quad \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^{\mathsf{T}} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^{\mathsf{T}} \boldsymbol{\mu}_{ij}$$

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

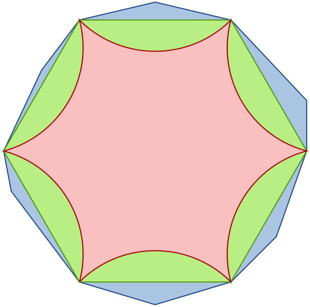
outer approximation

LP for Local Marginal Polytope

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



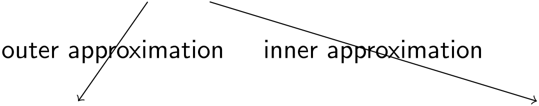
Linear and Quadratic Programming Relaxations



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

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

Quadratic Programming

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

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

LPQP for MAP Inference

LPQP: Best of Both Worlds?

Combine Linear and Quadratic Programming relaxations.

Joint LP and QP Objective

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{\mu} + \rho g(\boldsymbol{\mu}).$$

- $g(\boldsymbol{\mu})$: Penalty term that discourages discrepancy between

$$\boldsymbol{\mu}_{ij} \Leftrightarrow \boldsymbol{\mu}_i \boldsymbol{\mu}_j^{\mathsf{T}}$$

- Constraint set: the local marginal polytope $\mathcal{L}_{\mathcal{G}}$
- Non-convex
- Limit cases w.r.t ρ. LP relaxation: $\rho = 0$. QP relaxation: $\rho \rightarrow \infty$.

KL-Divergence Penalty Terms

Two Variants Of The Penalty Term

Uniform Weighting Of Edges:

$$\sum_{(i,j)\in\mathcal{E}} D_{KL}(\boldsymbol{\mu}_{ij}, \boldsymbol{\mu}_i \boldsymbol{\mu}_j^{\mathsf{T}})$$

Tree-based Weighting with $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a)$:

$$\sum_{a\in\mathcal{A}} \eta_a \left(\sum_{(i,j)\in\mathcal{E}_a} D_{KL}(\boldsymbol{\mu}_{ij}, \boldsymbol{\mu}_i \boldsymbol{\mu}_j^{\mathsf{T}}) \right)$$

* D_{KL} denotes the Kullback-Leibler divergence

LPQP Algorithm Overview

```
Require: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \boldsymbol{\theta}$.
initialize $\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}$ uniform, $\rho = \rho_0$.
repeat
    $t = 0, \boldsymbol{\mu}^0 = \boldsymbol{\mu}$.
    repeat
        $\boldsymbol{\mu}^{t+1} = \operatorname*{argmin}_{\boldsymbol{\tau} \in \mathcal{L}_{\mathcal{G}}} u_\rho(\boldsymbol{\tau}) - \boldsymbol{\tau}^{\mathsf{T}} \nabla v_\rho(\boldsymbol{\mu}^t)$.
        $t = t + 1$.
    until $\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^{t-1}\|_2 \leq \epsilon_{\mathrm{dc}}$.
    $\boldsymbol{\mu} = \boldsymbol{\mu}^t$.
    increase $\rho$.
until $\|\boldsymbol{\mu} - \boldsymbol{\mu}^0\|_2 \leq \epsilon_\rho$.
return $\boldsymbol{\mu}$.
```

LPQP Algorithm Overview

```
Require: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \boldsymbol{\theta}$.
initialize $\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}$ uniform, $\rho = \rho_0$.
repeat
    $t = 0, \boldsymbol{\mu}^0 = \boldsymbol{\mu}$.
    repeat
        $\boldsymbol{\mu}^{t+1} = \operatorname*{argmin}_{\boldsymbol{\tau} \in \mathcal{L}_{\mathcal{G}}} u_\rho(\boldsymbol{\tau}) - \boldsymbol{\tau}^\mathsf{T} \nabla v_\rho(\boldsymbol{\mu}^t)$.
        $t = t + 1$.
    until $\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^{t-1}\|_2 \leq \epsilon_{\mathrm{dc}}$.
    $\boldsymbol{\mu} = \boldsymbol{\mu}^t$.
    increase $\rho$.
until $\|\boldsymbol{\mu} - \boldsymbol{\mu}^0\|_2 \leq \epsilon_\rho$.
return $\boldsymbol{\mu}$.
```

CCCP For a Given ρ

- Non-convex:
  apply concave-convex procedure.

LPQP Algorithm Overview

```
Require: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \boldsymbol{\theta}$.
initialize $\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}$ uniform, $\rho = \rho_0$.
repeat
    $t = 0, \boldsymbol{\mu}^0 = \boldsymbol{\mu}$.
    repeat
        $\boldsymbol{\mu}^{t+1} = \underset{\boldsymbol{\tau} \in \mathcal{L}_{\mathcal{G}}}{\text{argmin}}\, u_\rho(\boldsymbol{\tau}) - \boldsymbol{\tau}^{\mathsf{T}} \nabla v_\rho(\boldsymbol{\mu}^t)$.
        $t = t + 1$.
    until $\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^{t-1}\|_2 \leq \epsilon_{\mathrm{dc}}$.
    $\boldsymbol{\mu} = \boldsymbol{\mu}^t$.
    $\boxed{\text{increase } \rho.}$
until $\|\boldsymbol{\mu} - \boldsymbol{\mu}^0\|_2 \leq \epsilon_\rho$.
return $\boldsymbol{\mu}$.
```

Gradual Increment of ρ

- Gradually moving from an LP solution to a QP solution.

LPQP Algorithm Overview

```
Require: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \boldsymbol{\theta}$.
initialize $\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}$ uniform, $\rho = \rho_0$.
repeat
    $t = 0, \boldsymbol{\mu}^0 = \boldsymbol{\mu}$.
    repeat
        $\boldsymbol{\mu}^{t+1} = \boxed{\operatorname*{argmin}_{\boldsymbol{\tau} \in \mathcal{L}_{\mathcal{G}}} u_\rho(\boldsymbol{\tau}) - \boldsymbol{\tau}^\mathsf{T} \nabla v_\rho(\boldsymbol{\mu}^t).}$
        $t = t + 1$.
    until $\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^{t-1}\|_2 \leq \epsilon_{\mathrm{dc}}$.
    $\boldsymbol{\mu} = \boldsymbol{\mu}^t$.
    increase $\rho$.
until $\|\boldsymbol{\mu} - \boldsymbol{\mu}^0\|_2 \leq \epsilon_\rho$.
return $\boldsymbol{\mu}$.
```

Convex Sub-Problems

Have the form:

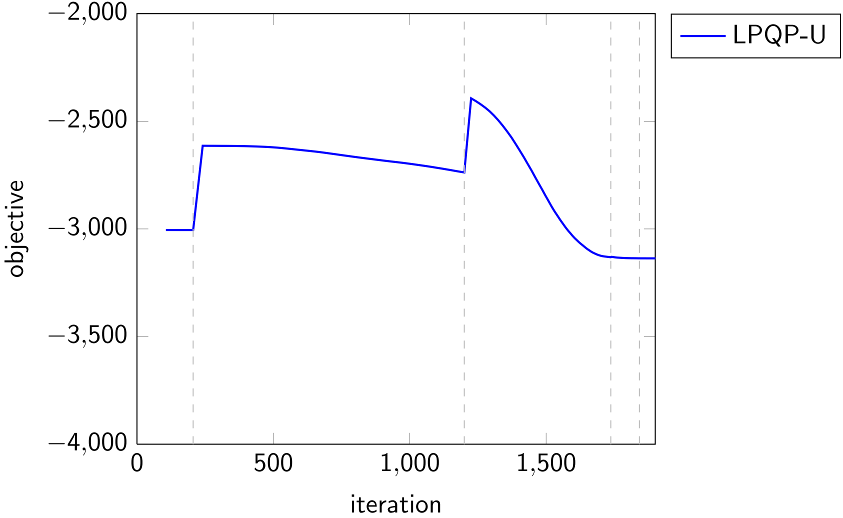
$$\min_{\boldsymbol{\mu} \in \mathcal{L}_{\mathcal{G}}} \mathrm{LP}(\boldsymbol{\mu})^1 - \rho \cdot \mathrm{Entropy}(\boldsymbol{\mu})^2$$

① LP with modified unary potentials → $\tilde{\boldsymbol{\theta}}$

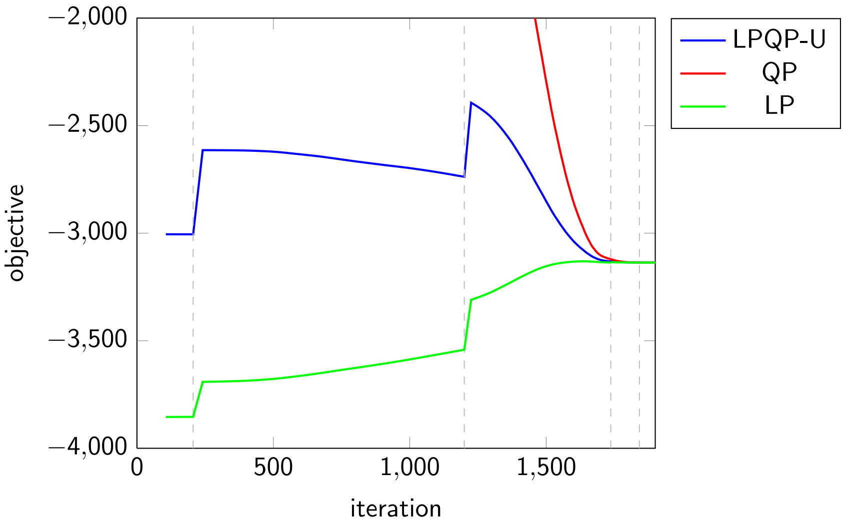
② Entropy term, different between the penalty variants

Solved efficiently with message passing algorithms

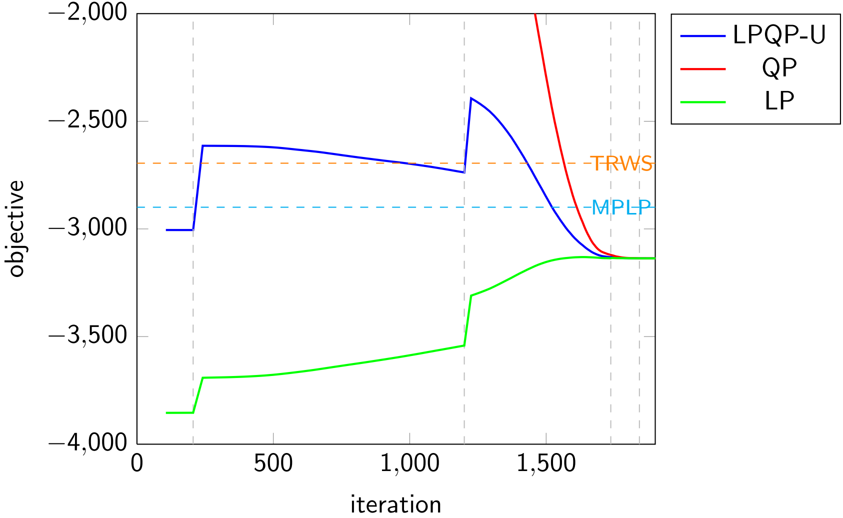
A Run of LPQP


−2,000
−2,500
−3,000
−3,500
−4,000
objective
0
500
1,000
1,500
iteration
LPQP-U

A Run of LPQP


−2,000
−2,500
−3,000
−3,500
−4,000
objective
0
500
1,000
1,500
iteration
LPQP-U
QP
LP

A Run of LPQP


−2,000
−2,500
−3,000
−3,500
−4,000
0
500
1,000
1,500
objective
iteration
TRWS
MPLP
LPQP-U
QP
LP