LPQP for MAP: Putting LP Solvers to Better Use

Patrick Pletscher, Sharon Wulff

Machine Learning Laboratory, ETH Zürich, Switzerland

ICML
June 28, 2012
### Energy Minimization

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

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

For general $\mathcal{G}$ and $\theta$ this is an NP-hard problem.
Linear and Quadratic Programming Relaxations

**LP for Marginal Polytope**

$$\min_{\mu \in \mathcal{M}_G} \sum_{i \in \mathcal{V}} \theta_i^T \mu_i + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}^T \mu_{ij}$$

But: $\mathcal{M}_G$ is exponentially large!
Linear and Quadratic Programming Relaxations

LP for Marginal Polytope

\[
\min_{\mu \in \mathcal{M}_G} \sum_{i \in \mathcal{V}} \theta_i^T \mu_i + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}^T \mu_{ij}
\]

But: \( \mathcal{M}_G \) is exponentially large!

LP for Local Marginal Polytope

\[
\min_{\mu \in \mathcal{L}_G} \sum_{i \in \mathcal{V}} \theta_i^T \mu_i + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}^T \mu_{ij}
\]
Linear and Quadratic Programming Relaxations

LP for Marginal Polytope

\[ \min_{\mu \in \mathcal{M}_G} \sum_{i \in V} \theta_i^T \mu_i + \sum_{(i,j) \in E} \theta_{ij}^T \mu_{ij} \]

But: \( \mathcal{M}_G \) is exponentially large!

outer approximation  inner approximation

LP for Local Marginal Polytope

\[ \min_{\mu \in \mathcal{L}_G} \sum_{i \in V} \theta_i^T \mu_i + \sum_{(i,j) \in E} \theta_{ij}^T \mu_{ij} \]

Quadratic Programming

\[ \min_{\mu \in \mathcal{L}_G} \sum_{i \in V} \theta_i^T \mu_i + \sum_{(i,j) \in E} \theta_{ij}^T \mu_{ij} \]

s.t. \( \mu_{ij} = \mu_i \mu_j^T \) \( \forall (i,j) \in E \)
LPQP for MAP Inference

LPQP: Best of Both Worlds?
Combine Linear and Quadratic Programming relaxations.

Joint LP and QP Objective

$$\min_{\mu \in \mathcal{L}_G} \theta^T \mu + \rho g(\mu).$$

- $g(\mu)$: Penalty term that discourages discrepancy between $\mu_{ij} \Leftrightarrow \mu_i^T \mu_j$

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

Patrick Pletscher, Sharon Wulff
KL-Divergence Penalty Terms

Two Variants Of The Penalty Term

Uniform Weighting Of Edges:

$$\sum_{(i,j) \in E} D_{KL}(\mu_{ij}, \mu_i \mu_j^T)$$

Tree-based Weighting with $G_a = (V_a, E_a)$:

$$\sum_{a \in A} \eta_a \left( \sum_{(i,j) \in E_a} D_{KL}(\mu_{ij}, \mu_i \mu_j^T) \right)$$

* $D_{KL}$ denotes the Kullback-Leibler divergence
LPQP Algorithm Overview

Require: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \theta$.

1: initialize $\mu \in \mathcal{L}_G$ uniform, $\rho = \rho_0$.

2: repeat
3: $t = 0, \mu^0 = \mu$.

4: repeat
5: $\mu^{t+1} = \arg\min_{\tau \in \mathcal{L}_G} u_\rho(\tau) - \tau^T \nabla v_\rho(\mu^t)$.
6: $t = t + 1$.

7: until $\|\mu^t - \mu^{t-1}\|_2 \leq \epsilon_{dc}$.

8: $\mu = \mu^t$.

9: increase $\rho$.

10: until $\|\mu - \mu^0\|_2 \leq \epsilon_\rho$.

11: return $\mu$. 
LPQP Algorithm Overview

Require: \( G = (\mathcal{V}, \mathcal{E}), \theta \).

1: initialize \( \mu \in \mathcal{L}_G \) uniform, \( \rho = \rho_0 \).

2: repeat
3: \( t = 0, \mu^0 = \mu \).

4: repeat
5: \( \mu^{t+1} = \arg\min_{\tau \in \mathcal{L}_G} u_\rho(\tau) - \tau^T \nabla \psi_\rho(\mu^t) \).
6: \( t = t + 1 \).
7: until \( \|\mu^t - \mu^{t-1}\|_2 \leq \epsilon_{dc} \).

8: \( \mu = \mu^t \).
9: increase \( \rho \).
10: until \( \|\mu - \mu^0\|_2 \leq \epsilon_\rho \).
11: return \( \mu \).

CCCP For a Given \( \rho \)

- Non-convex:
  apply concave-convex procedure.
LPQP Algorithm Overview

Require: $G = (V, E), \theta$.

1: initialize $\mu \in L_G$ uniform, $\rho = \rho_0$.
2: repeat
3: $t = 0, \mu^0 = \mu$.
4: repeat
5: $\mu^{t+1} = \arg\min_{\tau \in L_G} u_\rho(\tau) - \tau^T \nabla v_\rho(\mu^t)$.
6: $t = t + 1$.
7: until $\|\mu^t - \mu^{t-1}\|_2 \leq \epsilon_{dc}$.
8: $\mu = \mu^t$.
9: increase $\rho$.
10: until $\|\mu - \mu^0\|_2 \leq \epsilon_\rho$.
11: return $\mu$.

Gradual Increment of $\rho$

- Gradually moving from an LP solution to a QP solution.
LPQP Algorithm Overview

**Require:** \( G = (\mathcal{V}, \mathcal{E}), \theta. \)

1: initialize \( \mu \in \mathcal{L}_G \) uniform, \( \rho = \rho_0. \)

2: repeat
3: \( t = 0, \mu^0 = \mu. \)
4: repeat
5: \[ \mu^{t+1} = \underset{\tau \in \mathcal{L}_G}{\text{argmin}} u_\rho(\tau) - \tau^T \nabla v_\rho(\mu^t). \]
6: \( t = t + 1. \)
7: until \( \|\mu^t - \mu^{t-1}\|_2 \leq \epsilon_{dc}. \)
8: \( \mu = \mu^t. \)
9: increase \( \rho. \)
10: until \( \|\mu - \mu^0\|_2 \leq \epsilon_\rho. \)
11: return \( \mu. \)

**Convex Sub-Problems**

Have the form:

\[
\min_{\mu \in \mathcal{L}_G} \text{LP}(\mu)^1 - \rho \cdot \text{Entropy}(\mu)^2
\]

1. LP with modified unary potentials \( \rightarrow \tilde{\theta} \)
2. Entropy term, different between the penalty variants

Solved efficiently with message passing algorithms
A Run of LPQP

The graph shows the objective function over iterations for LPQP-U. The x-axis represents the iteration number, ranging from 0 to 1,500, while the y-axis represents the objective function values, ranging from -4,000 to -2,000.
A Run of LPQP

The diagram shows the objective function over iterations for three methods:
- **LPQP-U** (blue line)
- **QP** (red line)
- **LP** (green line)

The x-axis represents the iteration number, and the y-axis represents the objective value. The graph illustrates how each method converges to different solutions.
A Run of LPQP