Solving Markov Random Fields with Spectral Relaxation

Timothee Cour Jianbo Shi

GRASP — University of Pennsylvania

Many problems in computer vision and Machine Learning can be formulated using Markov Random Fields

proposed approach:

- **no restrictions** on the pairwise clique potentials
- **complexity O(#edges #labels^2),** ***linear*** **in the description length of the clique potentials**
- **improved general optimality bounds**

MRFs and IQP/QP formulation

$$\max \; P(X) = \frac{1}{Z} \prod_{ij \in E} \Psi_{ij}(X_i, X_j) \prod_i \Phi_i(X_i)$$

$$X_i \in \{1, ..., k\}$$

IQP formulation

$x_{ia} = 1$ iff $X_i = a$ $\quad \epsilon(x) = x^\top W x + V^\top x$

$\max \epsilon(x)$ s.t. $Cx = 1, x \in \{0,1\}^{nk}$ — discrete set Ω_d

QP relaxation

$\max \epsilon(x)$ s.t. $Cx = 1, 0 \le x \le 1$ — simplex Ω_s

non-convex !

Proposition: QP is equivalent to IQP

(generalizes a result form Ravikumar & Lafferty, 2006)
given a solution to IQP, we can construct a solution to QP

Solving Affine Constrained Rayleigh Quotients

$$\max_x \frac{x^\top W x + V^\top x}{x^\top x + \beta} \quad \text{s.t.} \quad Cx = b$$

EQUIVALENT to IQP for x binary

Linear Constraint: $Cx = 0$ (and $V = 0, \beta = 0$) — Yu and Shi, 2001

Affine Constraint: $Cx = b \iff \sum_a x_{ia} = 1 \implies$ SQP

Inequality Constraint ? $Cx \le b$ — Theorem: this is NP-HARD

Solution

1. rewrite as $\max_{x,t} \frac{x^\top W x + t V^\top x}{x^\top x + \beta t^2}$ s.t. $Cx = tb$ then $\max \frac{\bar{x}^\top \bar{W} \bar{x}}{\bar{x}^\top \bar{D} \bar{x}}$ s.t. $\bar{C}\bar{x} = 0$
2. introduce $x' = \bar{D}^{1/2}\bar{x} \implies \max \epsilon_1(x') = \frac{x'^\top W' x'}{x'^\top x'}$ s.t. $C'x' = 0$
3. solve $P_C W' P_C \quad x' = \lambda x'$ with $P_C = I - C'^\top (C'C'^\top)^{-1} C'$

Efficient computation with Shermann-Morrison formula

Complexity (per eigensolver iteration): $O(nnz(W')) = O(nnz(W)) = O(k^2|E|)$

Spectral Relaxation to the QP (SQP)

$$\max \epsilon_S(x) = \frac{x^\top W x + V^\top x}{x^\top x + \beta} \quad \text{s.t.} \quad Cx = 1$$

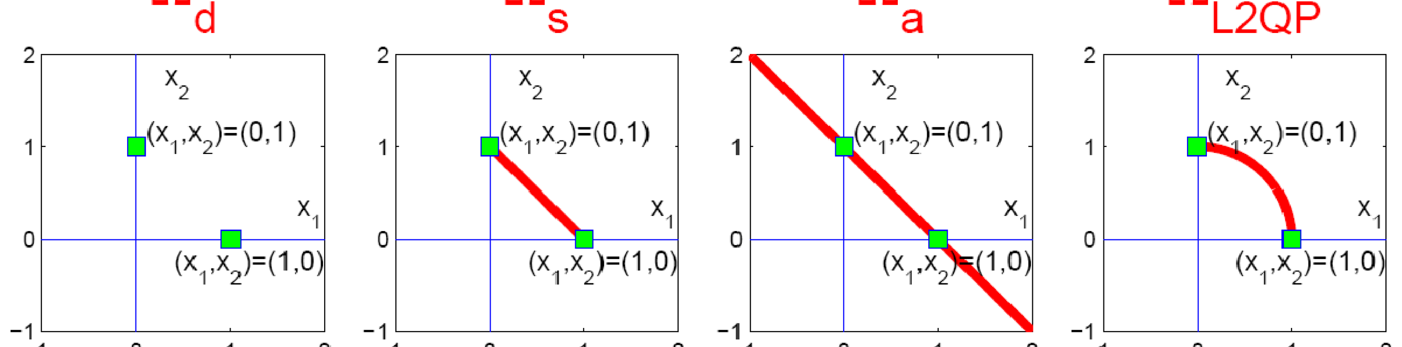
affine space Ω_a

normalization encourages ||x|| small, and $x \in [0, 1]$

still non-convex ! but solvable

different constraints

Ω_d $\quad \Omega_s$ $\quad \Omega_a$ $\quad \Omega_{L2QP}$



Algorithm

1. Input: clique potentials W,V
2. Set $\beta = \hat{\beta}$
3. Compute x_S from x', first eigenvector of $W_{eq} = P_C W' P_C$
4. Output upper bound $\epsilon^* \le \frac{n+\beta}{x_S^\top x_S + \beta} \epsilon(x_S)$
5. Discretize using Relaxation Labeling $\implies x_d$
6. Output lower bound $\epsilon^* \ge \epsilon(x_d)$

upper/lower bounds

General upper/lower bounds

$\forall x \in \Omega_d, \quad \frac{1}{n+\beta}\epsilon(x) \quad = \epsilon_S(x)$

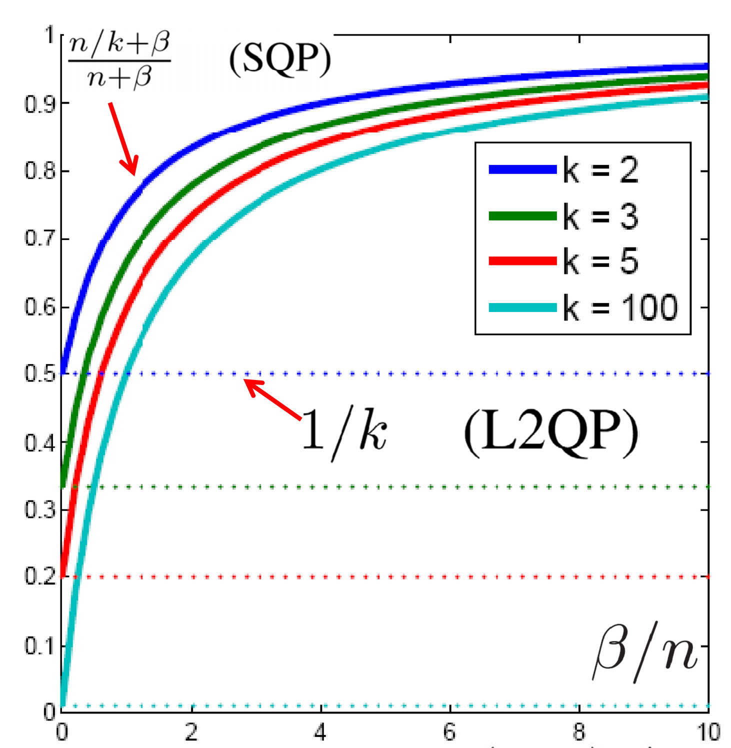
$\forall x \in \Omega_s, \quad \frac{1}{n+\beta}\epsilon(x) \quad \le \epsilon_S(x) \le \frac{1}{n/k+\beta}\epsilon(x)$

$\forall x \in \Omega_a, \quad \epsilon_S(x) \le \frac{1}{n/k+\beta}\epsilon(x)$

$(x^\top x \ge n/k,$ etc...)

From IQP ⇔ QP and the above, we get:

when $x_S \ge 0$,



lower bounds on $\epsilon(x_S)/\epsilon^*$

$$\epsilon(x_S) \le \epsilon(x_d) \le \epsilon^* \le \frac{n+\beta}{n/k+\beta}\epsilon(x_S)$$

Data dependent lower bound

$$M = W + diag(V)$$

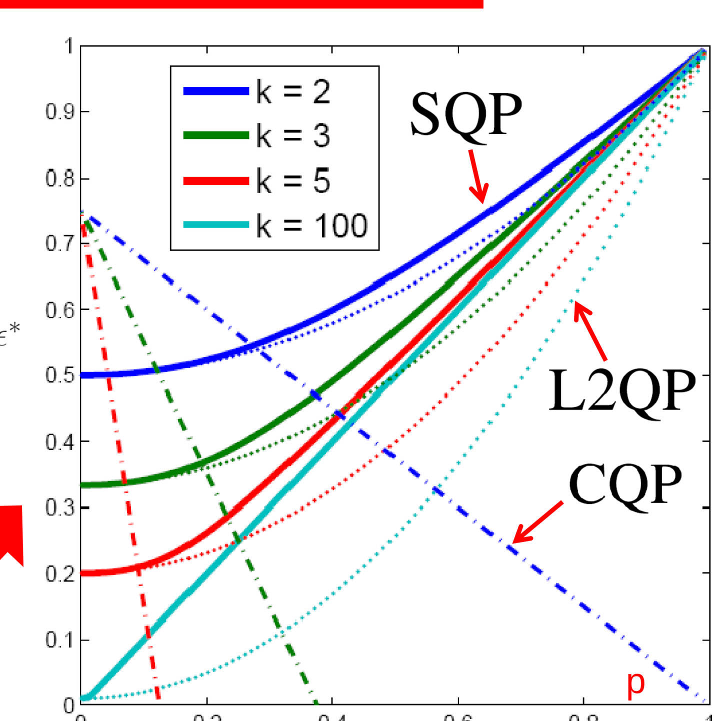
$\forall i, x^*_{i1} = 1$ and 0 otherwise $\quad \mathbf{1}^\top M_{1,1} \mathbf{1} = x^{*\top} M x^* = \epsilon^*$

$$M = \begin{bmatrix} M_{1,1} & M_{1,2:k} \\ M_{1,2:k}^\top & M_{2:k,2:k} \end{bmatrix}$$

p: the largest element in [0, 1] with

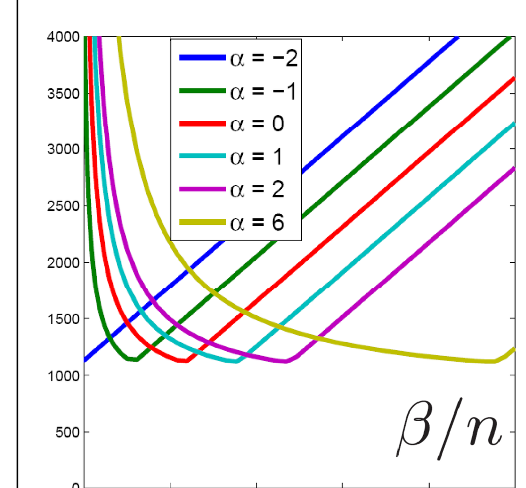
$pE[M_{1,1}] \le E[M_{2:k,2:k}]$
$pE[M_{1,1}] \le E[M_{1,2:k}]$ (peakedness)

$$\epsilon(x_s) \ge f(p, k)\epsilon^*$$



data-dependent lower bounds on $\epsilon(x_S)/\epsilon^*$ as function of p: peakedness of potentials

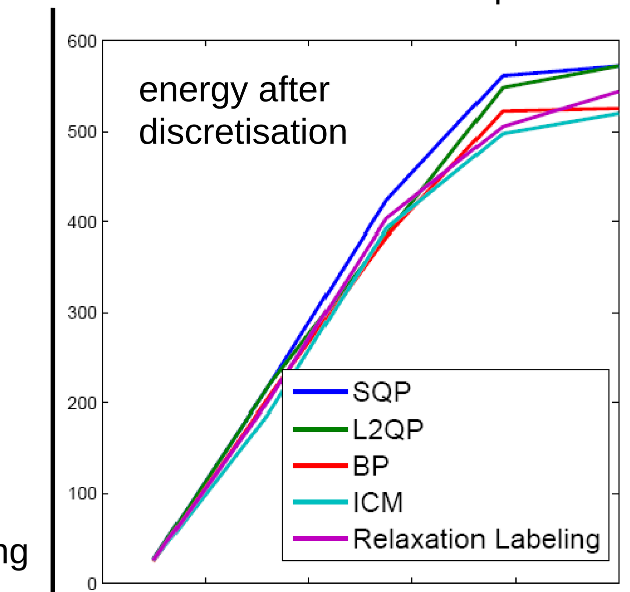
Experiments



family of upper bounds justifying heuristic:

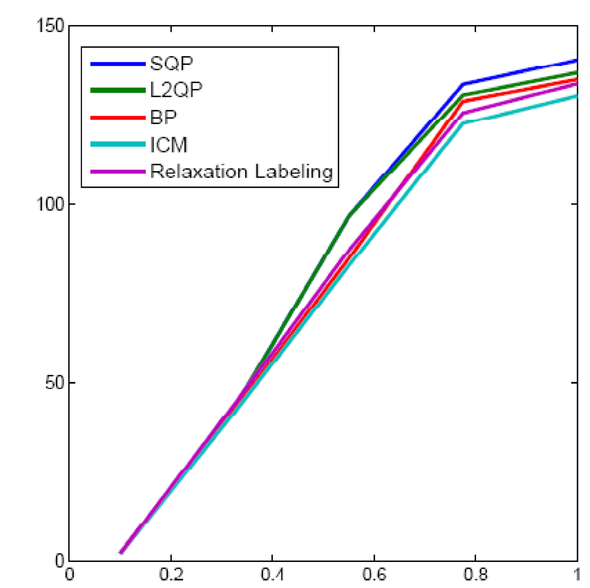
$$\beta_{opt} \approx \hat{\beta} = n^2 E[W] / \epsilon_S^{\bar{W},0*}$$

Comparison between SQP, L2QP, BP, ICM, Relaxation Labeling on random MRF problems with controlled parameters



density of connections

n=50, k=10



density of connections

n=20, k=20

Previous work

Linear relaxations: LP, SDP, SOCP

QP rewritten as (linear) matrix innter product, approximating rank 1 constraint

$$x^\top W x + V^\top x = \langle X, W_{eq} \rangle \text{ where } X = [x; 1][x; 1]^\top$$

Quadratic relaxations: L2QP and CQP

CQP: convexification of objective using (W,V)=> (W-diag(D),V+D)

L2QP: Cx=1 relaxed to $\sum_a x_{ia}^2 = 1$