Message-Passing for LP relaxations:

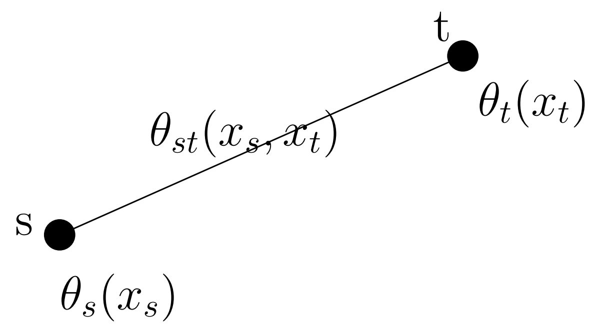
Proximal Bregman Projections with Convergence and Optimal Rounding Guarantees

Pradeep Ravikumar

(with Alekh Agarwal and Martin Wainwright)

University of California, Berkeley

Minimum Cost Labeling, *MAP*



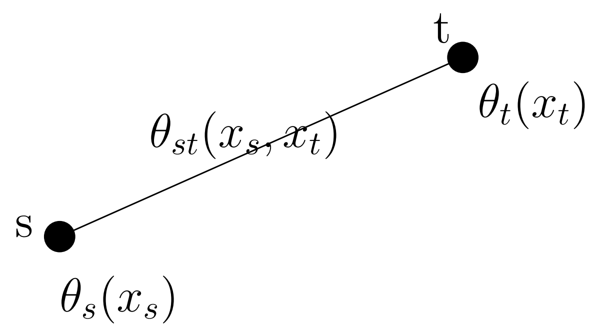
Label set : $L \ni x_s$

Node costs : $\theta_s(x_s)$

Edge costs : $\theta_{st}(x_s, x_t)$

$$\min_{x \in L^n} \sum_{s \in V(G)} \theta_s(x_s) + \sum_{(s,t) \in E(G)} \theta_{st}(x_s, x_t)$$

Minimum Cost Labeling, *MAP*



Label set : $L \ni x_s$

Node costs : $\theta_s(x_s)$

Edge costs : $\theta_{st}(x_s, x_t)$

$$\min_{x \in L^n} \sum_{s \in V(G)} \theta_s(x_s) + \sum_{(s,t) \in E(G)} \theta_{st}(x_s, x_t)$$

Linear Programming Relaxation

$$\min_{\mu \in \mathrm{LOCAL}(G)} \sum_{s} \sum_{x_s} \theta_s(x_s)\mu_s(x_s) + \sum_{(s,t)\in E} \sum_{x_s, x_t} \theta_{st}(x_s, x_t)\mu_{st}(x_s, x_t),$$

- Assign pseudomarginals μ_s and μ_{st} to nodes and edges instead of labels
- Pseudomarginals μ relax indicator variables.

Linear Programming Relaxation

$$\min_{\mu \in \mathrm{LOCAL}(G)} \sum_{s} \sum_{x_s} \theta_s(x_s)\mu_s(x_s) + \sum_{(s,t)\in E} \sum_{x_s,x_t} \theta_{st}(x_s,x_t)\mu_{st}(x_s,x_t),$$

$$\mathrm{LOCAL}(G) := \quad \begin{aligned} &\sum_{x_s} \mu_s(x_s) = 1 \\ &\sum_{x_s,x_t} \mu_{st}(x_s,x_t) = 1 \\ &\sum_{x_t} \mu_{st}(x_s,x_t) = \mu_s(x_s) \\ &\mu_s(x_s), \mu_{st}(x_s,x_t) \geq 0 \end{aligned}$$

(Schlesinger, 76), (Chekuri et al, 05), (KKT 07)

Graph-structured LPs

Simplex, Interior Point, Subgradient methods.

Iterative graph-structured algorithms?

Algorithms to solve graph-structured LPs?

Graph-structured message passing procedures:

- Each node collects info from neighbors
- Collates the info
- Distributes info to neighbors

Max product

$$M_{st}(x_t) \propto \max_{x_s} \left\{ \exp(\theta_{st}(x_s, x_t) \prod_{t \in N(s) \setminus t} M_{us}(x_s) \right\}$$

Max product

$$M_{st}(x_t) \propto \max_{x_s} \left\{ \exp(\theta_{st}(x_s, x_t) \prod_{t \in N(s) \setminus t} M_{us}(x_s) \right\}$$

- Need not converge
- Minimizes zero temperature limit of Bethe energy (not the LP)

Tree-reweighted Max Product

LP Relaxation:

$$\min_{\mu \in \mathrm{LOCAL}(G)} \sum_{s} \sum_{x_s} \theta_s(x_s)\mu_s(x_s) + \sum_{(s,t)\in E} \sum_{x_s,x_t} \theta_{st}(x_s,x_t)\mu_{st}(x_s,x_t),$$

Tree-reweighted Max Product

LP Relaxation:

$$\min_{\mu \in \mathrm{LOCAL}(G)} \sum_{s} \sum_{x_s} \theta_s(x_s)\mu_s(x_s) + \sum_{(s,t)\in E} \sum_{x_s, x_t} \theta_{st}(x_s, x_t)\mu_{st}(x_s, x_t),$$

Compact Notation:

$$\min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle$$

Tree-reweighted Max Product

LP:

$$\min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle$$

Dual:

$$\max_{\{\theta_T\}} \quad \sum_T \rho(T) \boxed{\text{MAP energy for tree } \theta_T}$$

$$\text{s.t.} \quad \sum_T \rho(T)\theta_T = \theta$$

"Dual decomposition" over trees.

[Wainwright et al, 2005]

Stopping Criterion, Rounding

$$\min_{\{\theta_T\}} \quad \sum_T \rho(T) \boxed{\text{MAP energy for tree } \theta_T}$$

$$\text{s.t.} \quad \sum_T \rho(T)\theta_T = \theta$$

Tree Agreement: At any iteration,

x^* optimal for all trees T in set $\mathcal{T} \Rightarrow x^*$ is the MAP solution.

[Wainwright et al, 2005]

Proximal Bregman Projections

Entropy Messages

Iterate, $n = 1, 2, \ldots$

Initialize:

$$\begin{aligned} \mu_{st}^{(n,0)} &= \mu_{st}^{(n-1)} \exp\left(w_n \ \theta_{st}\right) \\ \mu_{s}^{(n,0)} &= \mu_{s}^{(n-1)} \exp\left(w_n \ \theta_{s}\right) \end{aligned}$$

Iterate over edges (s,t):

$$\mu_{st}^{(n,\iota+1)}(x_s, x_t) \propto \mu_{st}^{(n,\iota)}(x_s, x_t) \sqrt{\frac{\mu_s^{(n,\iota)}(x_s)}{\sum_{x_t} \mu_{st}^{(n,\iota)}(x_s, x_t)}}$$

$$\mu_{s}^{(n,\iota+1)}(x_s) \propto \mu_{s}^{(n,\iota)}(x_s) \sqrt{\frac{\sum_{x_t} \mu_{st}^{(n,\iota)}(x_s, x_t)}{\mu_s^{(n,\iota)}(x_s)}}$$

Quadratic Messages

Iterate, $n = 1, 2, \ldots$

Intialize:

$$\begin{aligned} \mu_{st}^{(n,0)} &= \mu_{st}^{(n-1)} + w_n \ \theta_{st} \\ \mu_{s}^{(n,0)} &= \mu_{s}^{(n-1)} + w_n \ \theta_{s} \end{aligned}$$

Iterate over edges (s, t):

$$\begin{aligned} \mu_{st}^{(n,\iota+1)}(x_s, x_t) &= \mu_{st}^{(n,\iota)}(x_s, x_t) + \\ & (1/\ell + 1) \left(\mu_s^{(n,\iota)}(x_s) - \sum_{x_t} \mu_{st}^{(n,\iota)}(x_s, x_t) \right) \\ \mu_{s}^{(n,\iota+1)}(x_s) &= \mu_{s}^{(n,\iota)}(x_s) + \\ & (1/\ell + 1) \left(-\mu_s^{(n,\iota)}(x_s) + \sum_{x_t} \mu_{st}^{(n,\iota)}(x_s, x_t) \right) \end{aligned}$$

Proximal Iterations

$$LP: \quad \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle$$

Proximal Iterations

$$LP: \quad \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle$$

$$\mu^{(n+1)} = \arg \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle + w_n \, \|\mu - \mu^{(n)}\|^2$$

Proximal Bregman Iterations

$$LP: \quad \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle$$

$$\mu^{(n+1)} = \arg \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle + w_n \; D_f(\mu, \mu^{(n)})$$

(Bregman 67), (Censor and Zenios, 97)

Proximal Bregman Iterations

$$\begin{aligned}\mu^{(n+1)} &= \arg \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle + w_n D_f(\mu, \mu^{(n)}) \\ &= \arg \min_{\mu \in \mathrm{LOCAL}(G)} D_f(\mu, \widetilde{\mu^n})\end{aligned}$$

Updated parameters: $\widetilde{\mu^n} = \mu^{(n)} \odot \theta$

Entropy: $\widetilde{\mu^n} = \mu^{(n)} \exp(\theta)$

Quadratic: $\widetilde{\mu^n} = \mu^{(n)} + \theta$

Outline of the algorithm

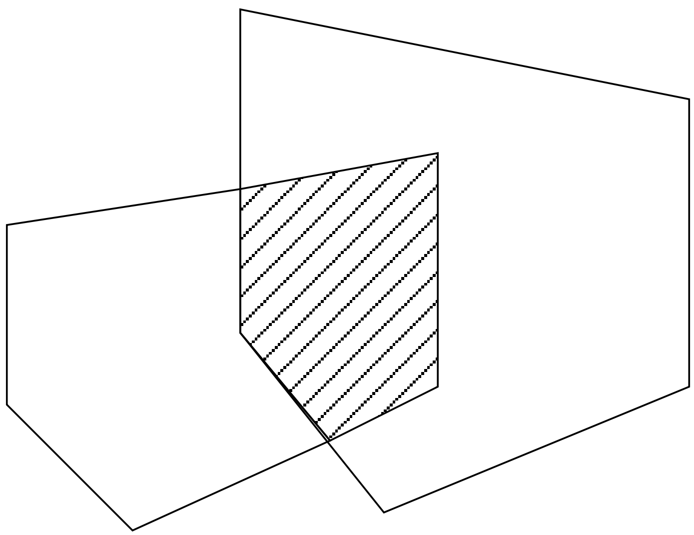
- Outer Loop:

$$\mu^{(n+1)} = \arg \min_{\mu \in \mathrm{LOCAL}(G)} \langle \theta, \mu \rangle + w_n D_f(\mu, \mu^{(n)})$$
$$= \arg \min_{\mu \in \mathrm{LOCAL}(G)} D_f(\mu, \widetilde{\mu^n})$$

  – Solve Outer loop by Inner loop:
    † Initialize: $\widetilde{\mu^n} = \mu^{(n)} \odot_f \theta$
    † $\mu^{(n+1)} \leftarrow$ Cyclic Bregman Projection of $\widetilde{\mu^n}$ onto local constraints.

Cyclic Projections

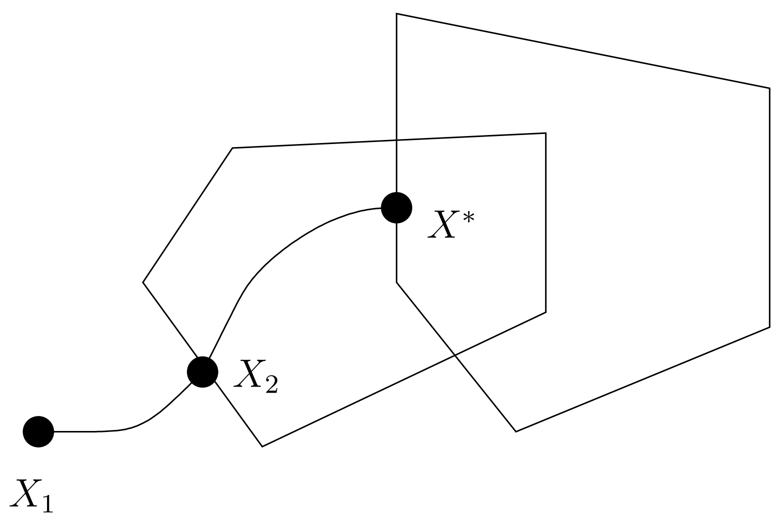
$$\min_{\mu \in \cap_i C_i} D_f(\mu, \widetilde{\mu})$$



[Bregman 67]

Cyclic Projections

$$\min_{\mu \in \cap_i C_i} D_f(\mu, \widetilde{\mu})$$



[Bregman 67]

Outline of the algorithm

- Outer Loop:

$$\mu^{(n+1)} = \arg \min_{\mu \in \text{LOCAL}(G)} \langle \theta, \mu \rangle + w_n D_f(\mu, \mu^{(n)})$$

  – Solve outer Loop by inner Loop:

  (a) Initialize: $\mu^{(n+1,0)} = \mu^{(n)} \odot_f \theta$

  (b) $\mu^{(n+1)} \leftarrow$ Cyclic Projection of $\widetilde{\mu^n}$.

    · For each edge $(s,t) \in E$,

    · $\mu^{(n+1,k+1)} \leftarrow$ Bregman Projection of $\mu^{(n+1,k)}$ onto (local) constraints for edge (s,t).

Entropy Messages

Iterate, $n = 1, 2, \ldots$

Intialize:

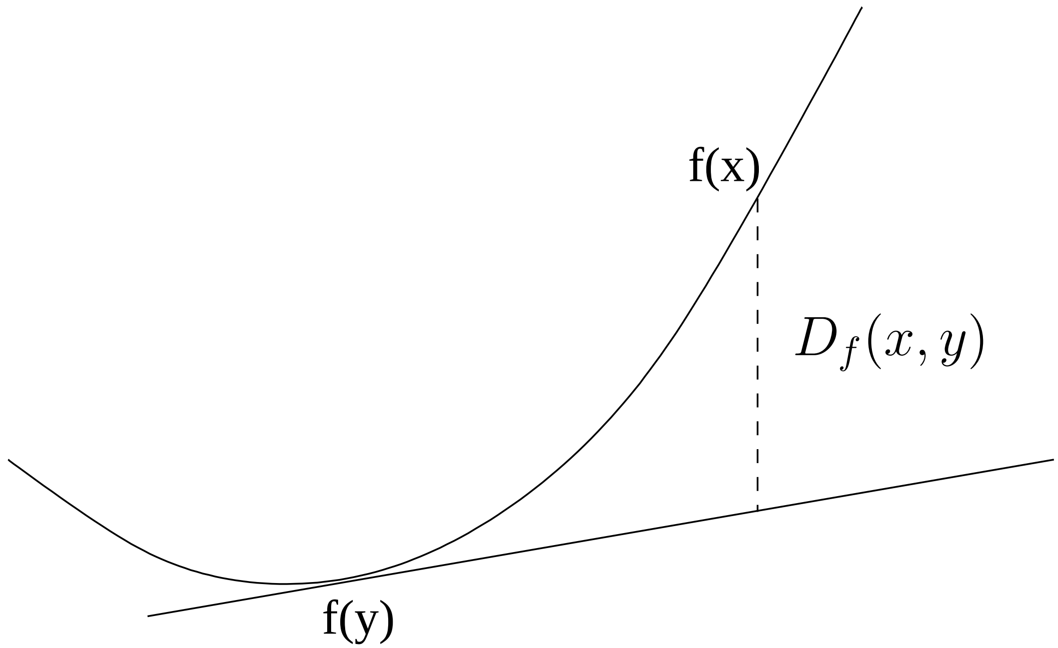
$$
\begin{aligned}
\mu_{st}^{(n,0)} &= \mu_{st}^{(n-1)} \exp\left(w_n \ \theta_{st}\right) \\
\mu_{s}^{(n,0)} &= \mu_{s}^{(n-1)} \exp\left(w_n \ \theta_{s}\right)
\end{aligned}
$$

Iterate over edges (s, t):

$$
\mu_{st}^{(n,\iota+1)}(x_s, x_t) \propto \mu_{st}^{(n,\iota)}(x_s, x_t) \sqrt{\frac{\mu_s^{(n,\iota)}(x_s)}{\sum_{x_t} \mu_{st}^{(n,\iota)}(x_s, x_t)}}
$$

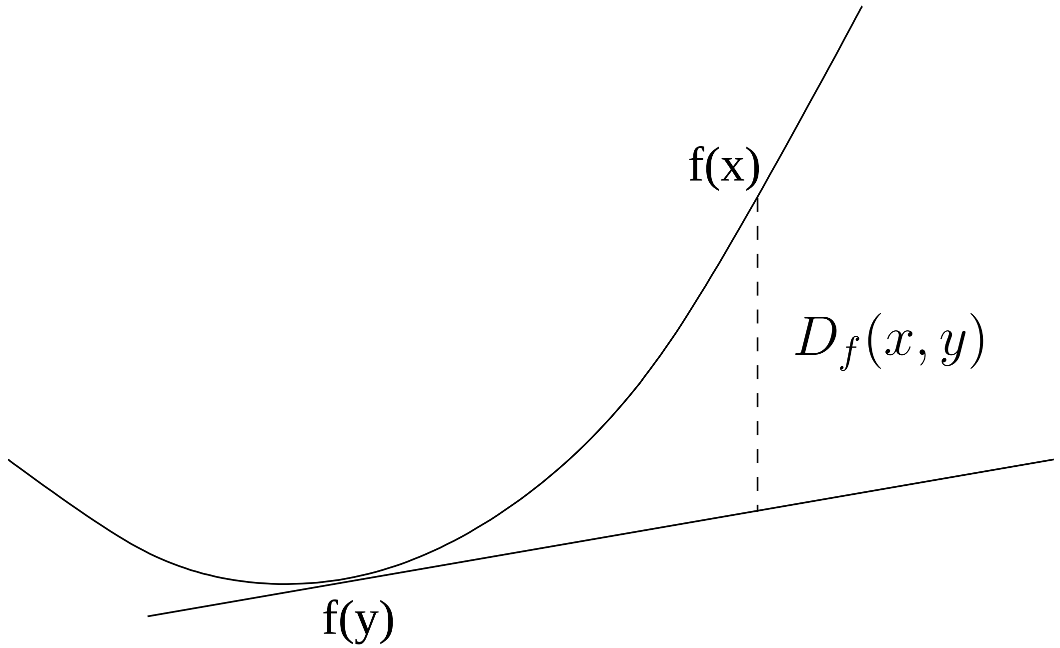
$$
\mu_{s}^{(n,\iota+1)}(x_s) \propto \mu_{s}^{(n,\iota)}(x_s) \sqrt{\frac{\sum_{x_t} \mu_{st}^{(n,\iota)}(x_s, x_t)}{\mu_s^{(n,\iota)}(x_s)}}
$$

Bregman Distances



$$D_f(x, y) = f(x) - f(y) - \nabla f(y)(x - y)$$

Bregman Distances



$$D_f(x, y) = f(x) - f(y) - \nabla f(y)(x - y)$$

f is convex: the more distant the point, the lower is the linear approximation bound.

Bregman Distances

$$D_f(y,x) = f(y) - f(x) - \nabla f(x)(y-x)$$

$$f(x) = \frac{1}{2}x^2 \quad : \quad D_f(x,y) = \frac{1}{2}\|x-y\|^2$$

$$f(x) = \sum_i x_i \log x_i \quad : \quad D_f(x,y) = \sum_i x_i \log(x_i/y_i) - x_i + y_i$$

Tree-reweighted Bethe Divergence:

$$f(\mu) \quad = \quad \sum_{s\in V, x_s} \rho_s \mu_s(x_s) \log \mu_s(x_s) + \sum_{(s,t)\in E, x_s, x_t} \rho_{st}\mu_{st}(x_s,x_t) \log \mu_{st}(x_s,x_t)$$

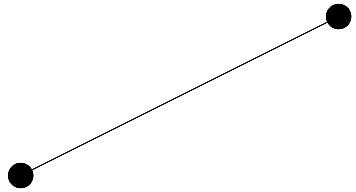
$$\Rightarrow \quad D_f(\mu,\nu) = \sum_{s\in V} D(\mu_s||\nu_s) + \sum_{(s,t)\in E} \rho_{st} E(\mu_{st}||\nu_{st})$$

Rounding Schemes

- Naive Rounding: $\widehat{x}_s = \arg\min_{x_s} \mu_s^n(x_s)$

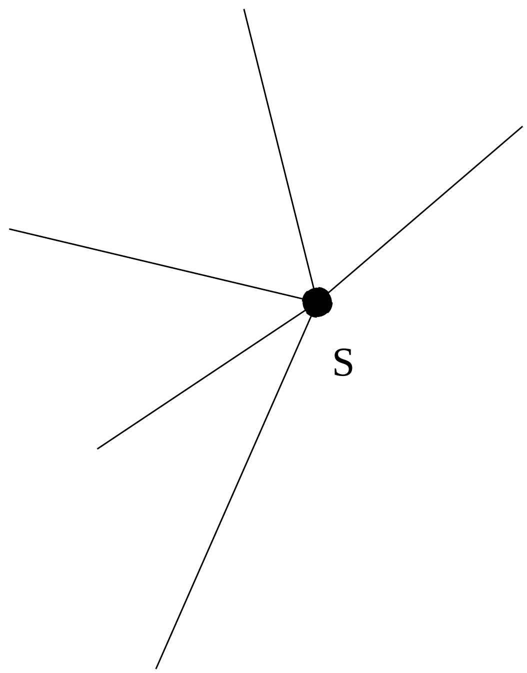
Graph based Rounding

- Naive Rounding
- Edge Rounding: $(\widehat{x}_s, \widehat{x}_t) = \arg\min_{x_s, x_t} \mu^n_{st}(x_s, x_t)$, *provided* they are consistent.



Graph based Rounding

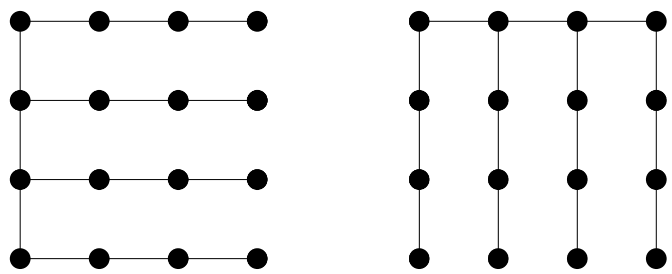
- Naive Rounding
- Edge Rounding
- Star Rounding: $\widehat{x} \leftarrow$ optimal configuration for each local neighborhood.



Graph based rounding

- Naive Rounding
- Edge Rounding
- Star Rounding

- Tree Rounding: $\widehat{x} \leftarrow$ optimal configuration for each tree in set.



Optimality Certificate

- Naive Rounding $\equiv$ no guarantees in general
- Theorem: At any iteration, if any of {Edge,Star,Tree} rounding schemes find a consistent configuration $\widehat{x}$, then $\widehat{x}$ is the MAP configuration.

Optimality Certificate

- Naive Rounding $\equiv$ no guarantees in general
- Theorem: At any iteration, if any of {Edge,Star,Tree} rounding schemes find a consistent configuration $\widehat{x}$, then $\widehat{x}$ is the MAP configuration.
- Computational Cost of Rounding: Naive $\leq$ Edge $\leq$ Star $\leq$ Tree

Order preserving Invariant

For the entropic messages:

$$\log\left\{\prod_{s}\mu_s^n(x_s)\prod_{(s,t)\in E(G)}\mu_{st}^n(x_s,x_t)\right\}\propto\left(\sum_{i=1}^{n}w_n\right)E(\theta,x)$$

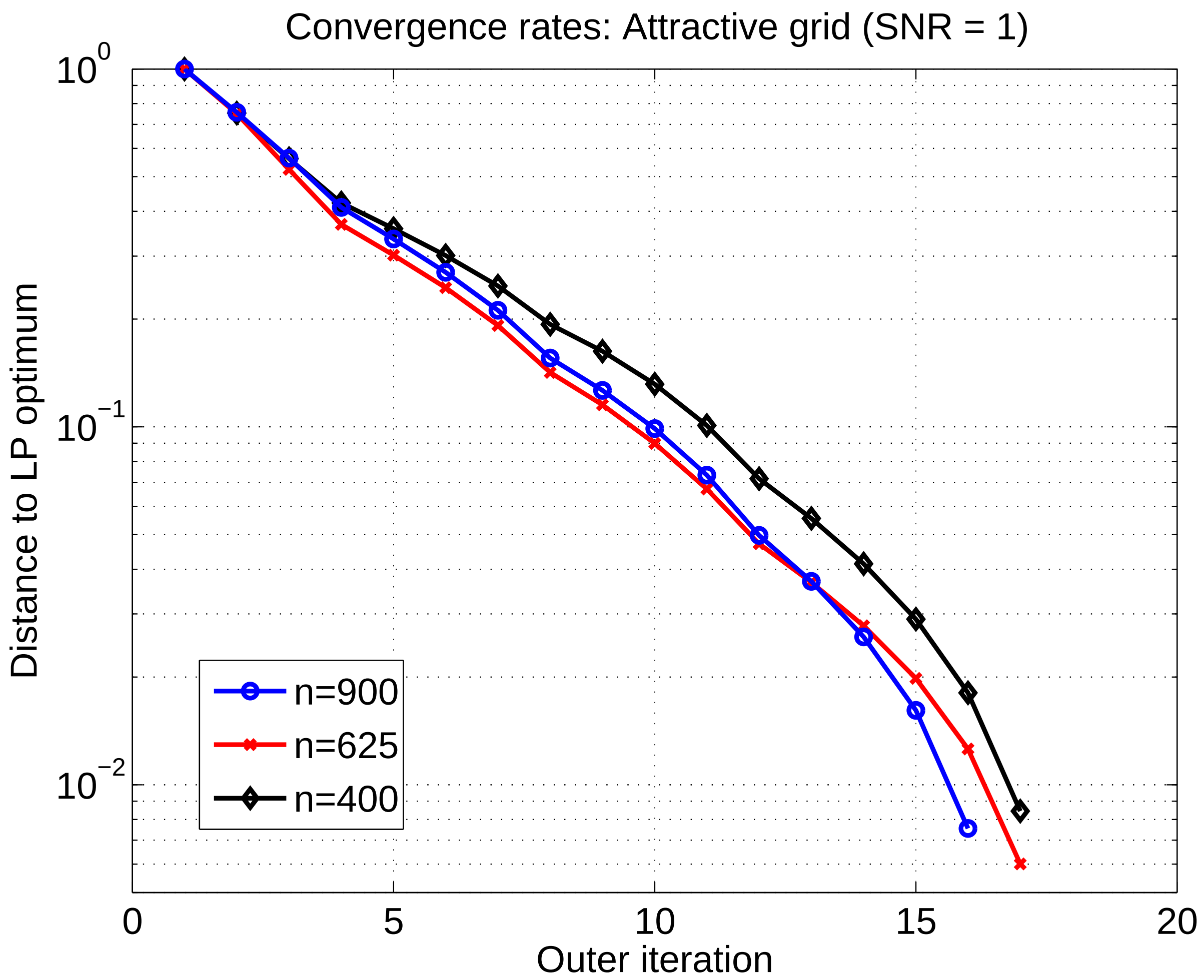
Convergence

Theorem: For any of the divergences {quadratic, entropic, TRW Bethe}, denoting the current iterate by μ^n and the LP optimum by μ^*, the convergence is super-linear

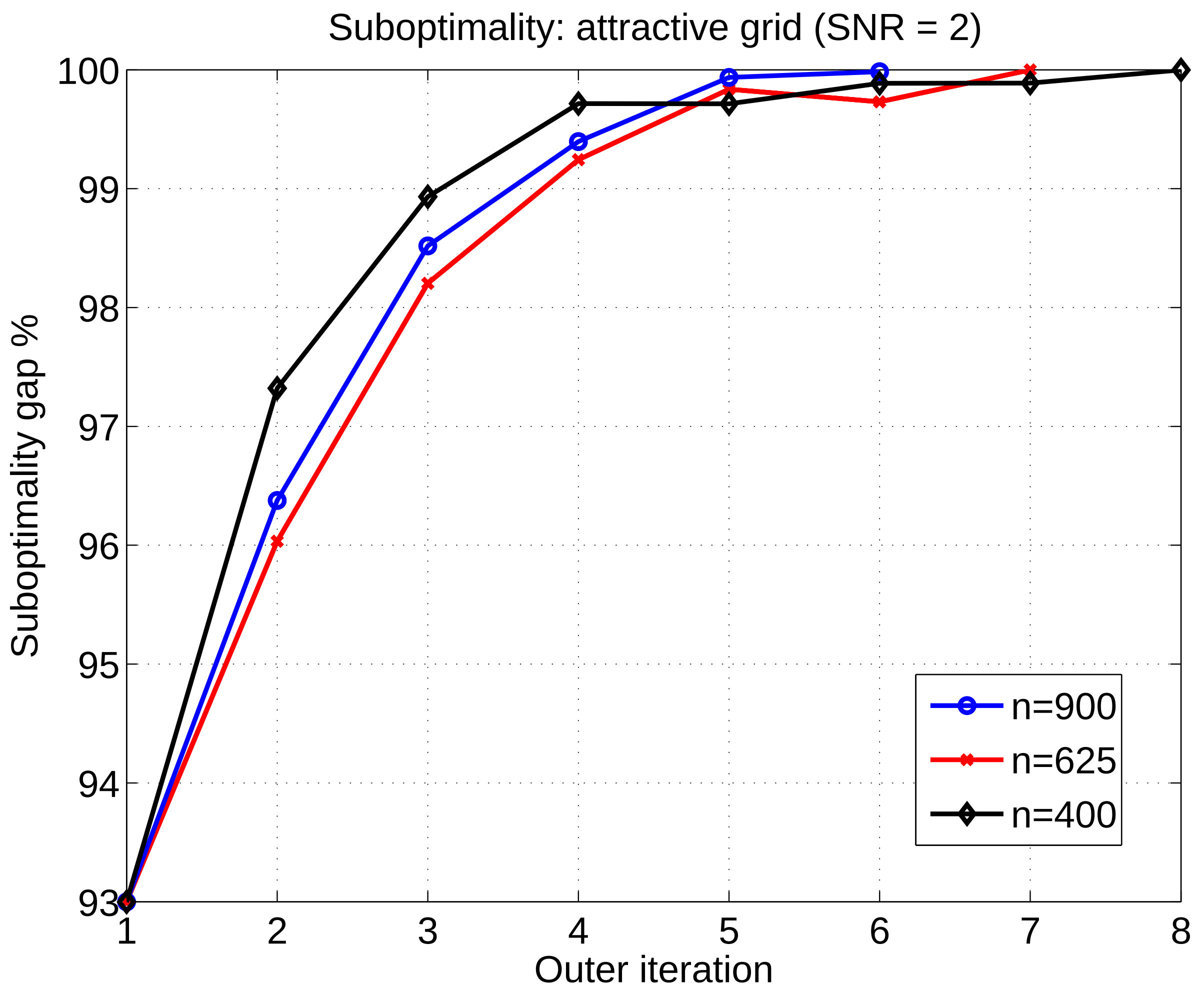
$$\lim_{n \to 0} \frac{\|\mu^n - \mu^*\|}{\|\mu^{n-1} - \mu^*\|} \to 0$$

Follows by adapting (Censor and Zenios, 97), (Teboulle, 95) and (Bertsekas and Tsitsiklis, 97)

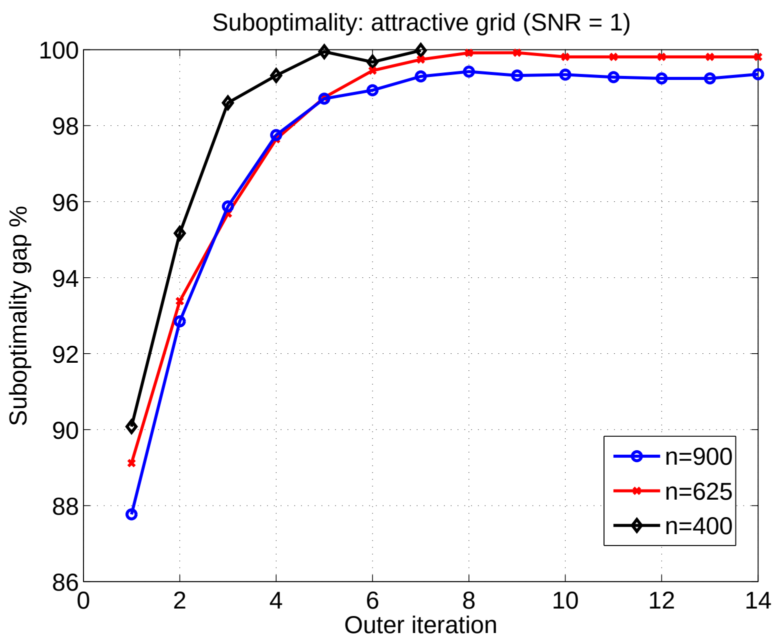
Convergence to LP optimum

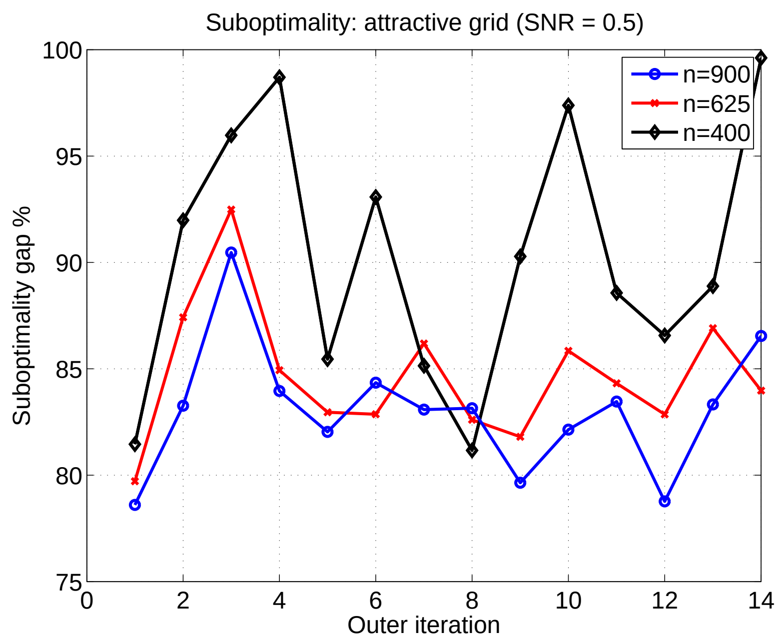


Optimality Gap

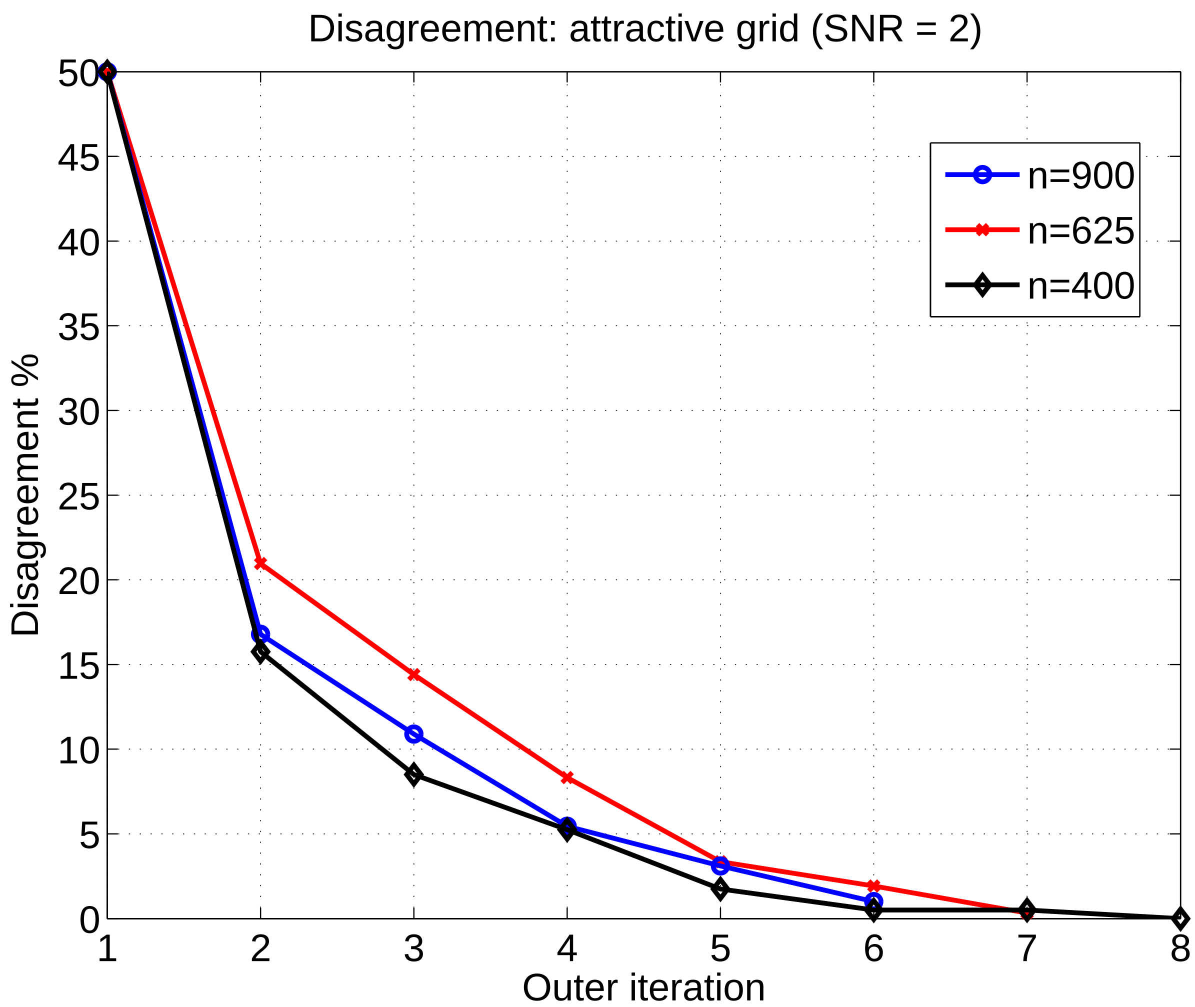


Optimality Gap





Tree Disagreement



Tree Disagreement