Efficient Energy Maximization Using Smoothing Technique

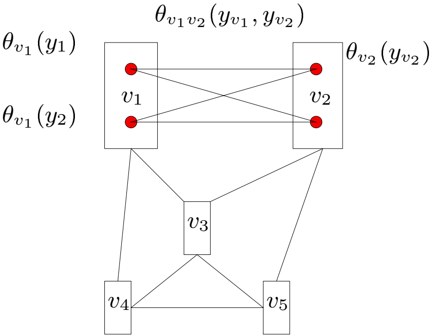
Bogdan Savchynskyy

Heidelberg Collaboratory for Image Processing (HCI)
University of Heidelberg

Energy Maximization Problem

$G = (\mathcal{V}, \mathcal{E}),\ v \in \mathcal{V},\ vv' \in \mathcal{E}$
$y_v \in \mathcal{Y}$ - *labels*, $y = (y_v, v \in \mathcal{V}) \in \mathcal{Y}^{\mathcal{V}} \equiv \mathcal{Y}^{|\mathcal{V}|}$
$\theta_t(y_v)$ - *unary potentials*, $\theta_{vv'}(y_v, y_{v'})$*-binary potentials*

$$y^* = \arg\max_{y \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} \theta_v(y_v) + \sum_{vv' \in \mathcal{E}} \theta_{vv'}(y_v, y_{v'}) \right] = \arg\max_{y \in \mathcal{Y}^{\mathcal{V}}} E(\theta, y)$$

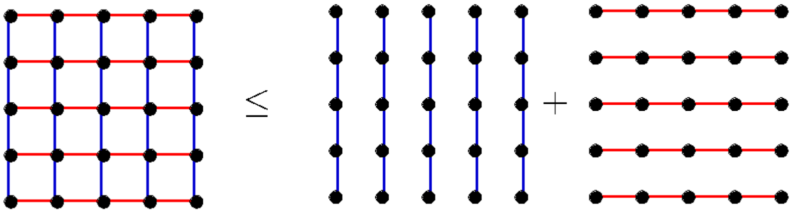


Dual Decomposition Approach to the Energy Maximization Problem

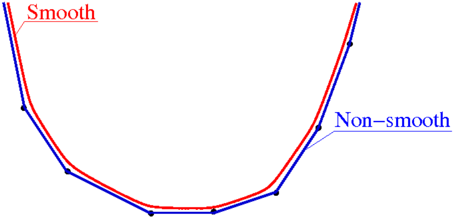
$$\theta = \theta^1 + \theta^2 \quad \Leftrightarrow \quad E(\theta, y) = E_1(\theta^1, y) + E_2(\theta^2, y)$$

$$\max_{y \in \mathcal{Y}^{\mathcal{V}}} E(\theta, y) \leqslant \max_{y \in \mathcal{Y}^{\mathcal{V}}} E_1(\theta^1, y) + \max_{y \in \mathcal{Y}^{\mathcal{V}}} E_2(\theta^2, y)$$

$$\max_{y \in \mathcal{Y}^{\mathcal{V}}} E(\theta, y) \leqslant \min_{\theta^1 + \theta^2 = \theta} \left[\max_{y \in \mathcal{Y}^{\mathcal{V}}} E_1(\theta^1, y) + \max_{y \in \mathcal{Y}^{\mathcal{V}}} E_2(\theta^2, y) \right]$$



Solving the Problem Efficiently

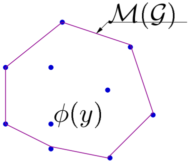


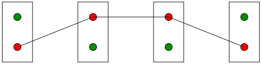
For our problem an effective and uniformly tight smoothing is possible.

Optimization method	**Convergence**
Non-smooth optimization (sub-gradient)	$O(1/\varepsilon^2)$
A smooth gradient descent	$O(\frac{L}{\varepsilon})$
↳ Applied to the non-smooth function	$L = \frac{1}{\varepsilon} \Rightarrow O(\frac{1}{\varepsilon^2})$
An optimal smooth first-order optimization:	$O(\sqrt{\frac{L}{\varepsilon}})$
↳ Applied to the non-smooth function	$L = \frac{1}{\varepsilon} \Rightarrow O(\frac{1}{\varepsilon})$

Outline

Energy Maximization Problem





$$\phi(y) = \begin{pmatrix} 01 & & 10 & & 10 & & 01 \\ & 0010 & & 1000 & & 0100 & \end{pmatrix}$$

$$\max_{y \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} \theta_v(y_v) + \sum_{vv' \in \mathcal{E}} \theta_{vv'}(y_v, y_{v'}) \right] \Rightarrow \max_{y \in \mathcal{Y}^{\mathcal{V}}} \sum_{c \in \mathcal{C}} \theta_c(y_c)$$

$$= \max_{y \in \mathcal{Y}^{\mathcal{V}}} \sum_{c \in \mathcal{C}} \langle \theta_c, \phi_c(y) \rangle = \max_{y \in \mathcal{Y}^{\mathcal{V}}} \langle \theta, \phi(y) \rangle = \max_{\mu \in \mathcal{M}(\mathcal{G})} \langle \theta, \mu \rangle$$

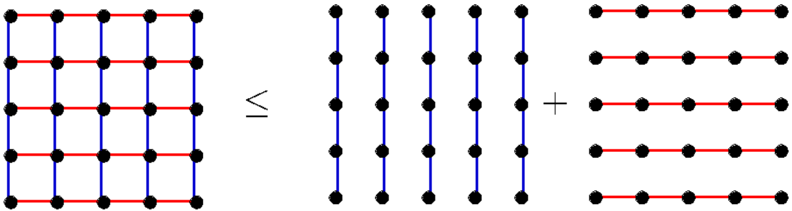
$$\phi_{(c,x')}(y) = \begin{cases} 1, & \text{if } y_c = x' \\ 0, & \text{otherwise} \end{cases}, \quad \mathcal{M}(\mathcal{G}) = \text{conv}\{\phi(y) : y \in \mathcal{Y}^{\mathcal{V}}\}$$

Decomposition

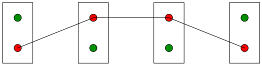
$$\theta = \theta^1 + \theta^2 \quad \Leftrightarrow \quad E(\theta, y) = E_1(\theta^1, y) + E_2(\theta^2, y)$$

$$\theta_c^1(\vartheta_c) = \begin{cases} \frac{\theta_c}{2} + \vartheta_c, & c \in \mathcal{V} \\ \theta_c, & c \in \mathcal{E}^1 \\ 0, & \text{otherwise} \end{cases} \qquad \theta_c^2(\vartheta_c) = \begin{cases} \frac{\theta_c}{2} - \vartheta_c, & c \in \mathcal{V} \\ \theta_c, & c \in \mathcal{E}^2 \\ 0, & \text{otherwise} \end{cases}$$

$$\max_{\mu \in \mathcal{M}(\mathcal{G})} \langle \theta, \mu \rangle \leqslant \max_{\mu^1 \in \mathcal{M}(\mathcal{G}^1)} \left\langle \theta^1(\vartheta), \mu^1 \right\rangle + \max_{\mu^2 \in \mathcal{M}(\mathcal{G}^2)} \left\langle \theta^2(\vartheta), \mu^2 \right\rangle$$



Smoothing



$$\phi(y) = (01 \quad 10 \quad 10 \quad 01)$$
$$0010 \quad 1000 \quad 0100$$

$$\max\{a_1, \ldots, a_n\} \simeq \rho \log\{e^{a_1/\rho} + \cdots + e^{a_n/\rho}\}$$

$$\max\{a_1, \ldots, a_n\} \leqslant \rho \log\{e^{a_1/\rho} + \cdots + e^{a_n/\rho}\} \leqslant \max\{a_1, \ldots, a_n\} + \rho \log n$$

$$U_{\mathcal{G}^i}(\vartheta) = \max_{\mu \in \mathcal{M}(\mathcal{G}^i)} \langle \theta^i(\vartheta), \mu \rangle = \max_{y \in \mathcal{Y}^{\mathcal{V}}} \langle \theta^i(\vartheta), \phi(y) \rangle$$

$$\hat{U}_{\mathcal{G}^i}(\vartheta, \rho) = \rho \log \sum_{y \in \mathcal{Y}^{\mathcal{V}}} \exp(\langle \theta^i(\vartheta)/\rho, \phi(y) \rangle)$$

$$U_{\mathcal{G}^i}(\vartheta) \leqslant \hat{U}_{\mathcal{G}^i}(\vartheta, \rho) \leqslant U_{\mathcal{G}^i}(\vartheta) + \rho \log |\mathcal{Y}^{\mathcal{V}}| = U_{\mathcal{G}^i}(\vartheta) + \rho |\mathcal{V}| \log |\mathcal{Y}|$$

$$\left(\frac{\partial \hat{U}_{\mathcal{G}^i}(\vartheta)}{\partial \vartheta_v(y_v)} \right) = \pm \rho \left(\frac{\sum_{y \in \mathcal{Y}(c, y_v)} \exp(\langle \theta^i(\vartheta)/\rho, \phi(y) \rangle)}{\hat{U}_{\mathcal{G}^i}(\theta^i(\vartheta), \rho))} \right)$$

Smoothing

$$U(\vartheta) = U_{\mathcal{G}^1}(\vartheta) + U_{\mathcal{G}^2}(\vartheta)\,.$$

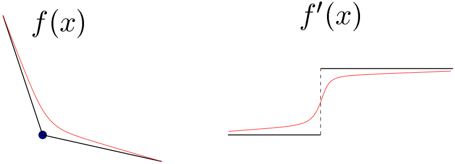
$$\hat{U}(\vartheta, \rho) = \hat{U}_{\mathcal{G}^1}(\vartheta, \rho) + \hat{U}_{\mathcal{G}^2}(\vartheta, \rho)$$

$$U(\vartheta) \leqslant \hat{U}(\vartheta, \rho) \leqslant U(\vartheta) + 2\rho|\mathcal{V}| \log|\mathcal{Y}|$$

$$\frac{\partial \hat{U}(\vartheta, \rho)}{\partial \vartheta} = \frac{\partial \hat{U}_{\mathcal{G}^1}(\vartheta, \rho)}{\partial \vartheta} + \frac{\partial \hat{U}_{\mathcal{G}^2}(\vartheta, \rho)}{\partial \vartheta}$$

Optimization: Lipschitz-Continuous Gradient

$f : \mathbb{R}^n \to \mathbb{R}$ – differentiable

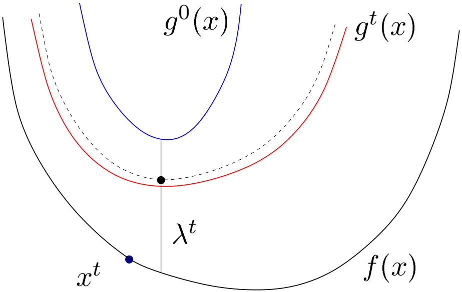


$$||\nabla f(x) - \nabla f(z)||_* \leqslant L||x - z||$$

Lemma (special case of Nesterov04)

Function $\hat{U}(\vartheta)$ is convex and Lipschitz-continuous with $L \leqslant \frac{2}{\rho}|\mathcal{V}|$

Optimization: Estimate Sequence [Nesterov83]



$$g^t(x) \leqslant (1-\lambda^t)f(x) + \lambda^t g^0(x), \quad \lambda^t \to 0$$
$$\{x^t\}: \quad f(x^t) \leqslant g^{t*} \equiv \min_{x\in\mathbb{R}^n} g^t(x)$$
$$f(x^t) - f^* \leqslant \lambda^t (g^0(x^*) - f^*) \to 0$$

Optimization: Algorithm

Algorithm (a variant of the algorithm 2.2.6 from Nesterov83)

$\gamma^t, \alpha^t, \omega^t, \in \mathbb{R}$; $x^t, u^t, z^t \in \mathbb{R}^n$, t – *an iteration counter.*

- *Choose* $x^0 \in \mathbb{R}^n$ *and* $\omega^0, \gamma^0 > 0$, $z^0 = u^0$.
- t*-th iteration* ($t \geqslant 0$).
  1. *Compute* $f(z^t)$ *and* $\nabla f(z^t)$.
  2. *Find possibly small* ω^t *such that*

$$f(x^t) \leqslant f(z^t) - \frac{1}{2\omega^t}||\nabla f(z^t)||^2 \text{ ,where } x^t = z^t - \frac{1}{\omega^t}\nabla f(z^t)\,.$$

  3. *Compute* $\alpha^t \in (0,1)$ *from the quadratic equation* $\omega^t(\alpha^t)^2 = (1-\alpha^t)\gamma^t$. *Set* $\gamma^{t+1} = (1-\alpha^t)\gamma^t$.
  4. *Set* $u^{t+1} = \frac{(1-\alpha^t)\gamma^t u^t - \alpha^t \nabla f(z^t)}{\gamma^{t+1}}$.
  5. *Choose* $z^{t+1} = \frac{\alpha^t \gamma^t u^{t+1} + \gamma^{t+1} x^t}{\gamma^t}$

Heidelberg Collaboratory for Image Processing

Optimization: Algorithm properties

Lemma (for any convex function with a Lipschitz-continuous gradient)

Condition at the step 2 in the algorithm is fulfilled for any $\omega^t \geqslant L$.

Lemma (modification of Nesterov83)

Convergence speed of the algorithm is $O(\frac{\sqrt{\omega^{*t}}}{t^2})$, *where* $\omega^{*t} = \max_{k \leqslant t} \omega^k$.

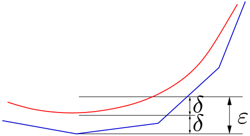
Corollary (modification of Nesterov83)

$\varepsilon = O(\frac{\sqrt{\omega^{*t}}}{t^2}) \Rightarrow t = O(\sqrt{\frac{\omega^{*t}}{\varepsilon}}); \quad \omega^{*t} \leqslant L = O(\frac{1}{\rho})$

Theorem (Algorithm convergence rate)

$$t = O\left(\sqrt{\frac{1}{\rho\varepsilon}}\right), \text{ if } \rho \sim \varepsilon \Rightarrow t = O\left(\frac{1}{\varepsilon}\right)$$

Smoothing selection



$$U(\vartheta) \leqslant \hat{U}(\vartheta, \rho) \leqslant U(\vartheta) + 2\rho|\mathcal{V}| \log|\mathcal{Y}|$$
$$\delta(\rho) = 2\rho|\mathcal{V}| \log|\mathcal{Y}|$$

[Nesterov04](the worst-case estimation) : $\rho = \frac{\varepsilon}{4|\mathcal{V}| \log|\mathcal{Y}|} \Rightarrow \delta(\rho) \leqslant \frac{\varepsilon}{2}$.

We estimate $\delta'(\rho) = \hat{U}(\vartheta, \rho) - U(\vartheta)$ for ϑ and use ρ such that $\delta'(\rho) \leqslant \frac{\varepsilon}{2}$.

Optimization: L Estimation

How large is $L = \frac{2}{\rho}|\mathcal{V}|$ for a typical setting?

$$\rho = 1,\ |\mathcal{V}| = 10 \times 10,\ \Rightarrow L = 2 \cdot 10^2$$

$$\rho = 1,\ |\mathcal{V}| = 100 \times 100,\ \Rightarrow L = 2 \cdot 10^4$$

Typical setting:

$$\rho = 1,\ |\mathcal{V}| = 400 \times 700,\ \Rightarrow L \approx 2 \cdot 10^5$$

We dynamically estimate L. Typical values are $1 - 1000$.

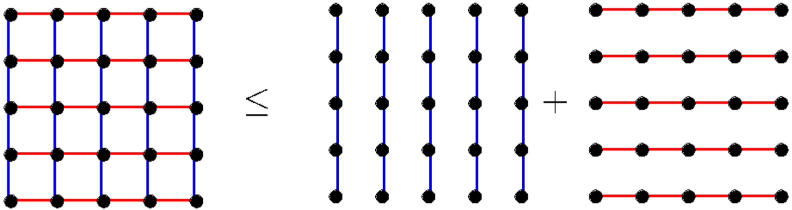
Optimization: ω^t Estimation

Algorithm (typical scheme of a linear search, implemented also in Nesterov05)

Input ω^t, $\nabla f(z^t)$, *parameters* $a > 1$ *and* $b > 1$, *output* ω^{t+1}, x^{t+1}.

1. *Set* $\omega^* = \omega^t / b$
2. *Calculate* $x^* = z^t - \frac{1}{\omega^*} \nabla f(z^t)$.
3. *If* $f(x^*) \leqslant f(z^t) - \frac{1}{2\omega^*} ||\nabla f(z^t)||^2$ **End.**
   Else assign $\omega^* = a \cdot \omega^*$ *goto step 2.*

Lower Bound: Standard Approach (covers only an LP-tight Case)



$\max\{E(\theta, y^1), E(\theta, y^2)\}$ $\leq$ y^1 $+$ y^2

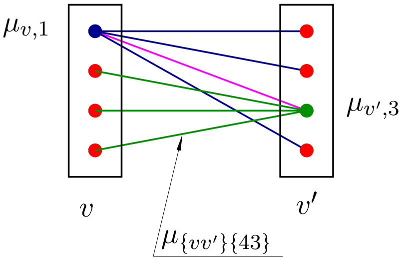
Lower Bound: Known Facts

$$\max_{\mu \in \mathcal{M}(\mathcal{G})} \langle \theta, \mu \rangle \leqslant \max_{\mu^1 \in \mathcal{M}(\mathcal{G}^1)} \left\langle \theta^1(\vartheta), \mu^1 \right\rangle + \max_{\mu^2 \in \mathcal{M}(\mathcal{G}^2)} \left\langle \theta^2(\vartheta), \mu^2 \right\rangle$$

if $\mathcal{G}^1$, $\mathcal{G}^2$ – trees, then:

$$\max_{\mu \in \mathcal{L}(\mathcal{G})} \langle \theta, \mu \rangle = \max_{\mu^1 \in \mathcal{M}(\mathcal{G}^1)} \left\langle \theta^1(\vartheta), \mu^1 \right\rangle + \max_{\mu^2 \in \mathcal{M}(\mathcal{G}^2)} \left\langle \theta^2(\vartheta), \mu^2 \right\rangle$$

where $\mathcal{L}(\mathcal{G})$ - a local polytope. Moreover, $\mathcal{M}(\mathcal{G}^i) = \mathcal{L}(\mathcal{G}^i)$.

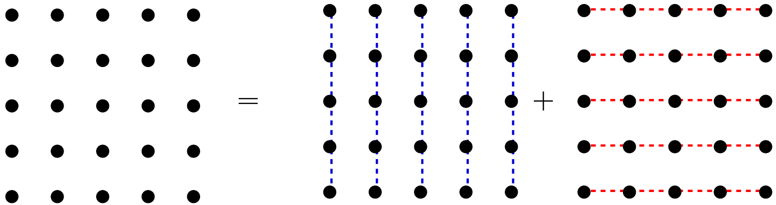


$$\begin{cases} \sum_{y' \in \mathcal{Y}} \mu_{\{vv'\},\{yy'\}} = \mu_{v,y} \\ \sum_{y \in \mathcal{Y}} \mu_{v,y} = 1 \\ \mu \geqslant 0 \end{cases}$$

Lower Bound: Our Approach (covers all cases)

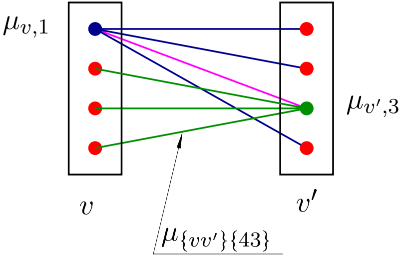
$$\mu_v^i = \pm\left(\frac{\partial \hat{U}_{\mathcal{G}^i}(\vartheta)}{\partial \vartheta_v(y_v)}\right) = \rho\left(\frac{\sum_{y\in\mathcal{Y}(c,y_v)} \exp(\langle \theta^i(\vartheta)/\rho, \phi(y)\rangle)}{\hat{U}_{\mathcal{G}^i}(\theta^i(\vartheta),\rho))}\right)$$

$\nabla\hat{U} = \mu_v^1 - \mu_v^2 \to 0$ close to the optima



$$\mu_v = \frac{\mu_v^1 + \mu_v^2}{2}$$

Lower Bound: Recovering $\mu_{vv'}$



$$\max \langle \theta, \mu \rangle$$

$$\begin{cases} \sum_{y' \in \mathcal{Y}} \mu_{\{vv'\},\{yy'\}} = \mu_{v,y} \\ \sum_{y \in \mathcal{Y}} \mu_{v,y} = 1 \\ \mu \geqslant 0 \end{cases}$$

hm... $\Downarrow$

$$\begin{cases} \sum_{y' \in \mathcal{Y}} \mu_{\{vv'\},\{yy'\}} = \mu_{v,y} \\ \text{for fixed } \mu_{v,y} = \frac{\mu^1_{v,y} + \mu^2_{v,y}}{2} \end{cases}$$

LP - Transportation problem!

Lower Bound: Theorem

Theorem

Let $\mu^{1,t} \in \mathcal{L}(\mathcal{G})$ *and* $\mu^{2,t} \in \mathcal{L}(\mathcal{G})$, $t = 1, \ldots \infty$ *be two sequences meeting the following conditions:*

1. $\mu_v^{1,t} - \mu_v^{2,t} \to 0,\ v \in \mathcal{V}$
2. $\left\langle \theta^i(\vartheta), \mu^{i,t}\big|_{\mathcal{L}(\mathcal{G}^i)} \right\rangle - \max_{\mu \in \mathcal{L}(\mathcal{G}^i)} \left\langle \theta^i(\vartheta), \mu \right\rangle \to 0,\ i = 1, 2$
3. $$\left\langle \theta, \frac{\mu^{1,t} + \mu^{2,t}}{2} \right\rangle = \max_{\mu \in \mathcal{L}(\mathcal{G})} \langle \theta, \mu \rangle \tag{1}$$
   $$\text{s.t. } \mu_v = \frac{\mu_v^{1,t} + \mu_v^{2,t}}{2},\ v \in \mathcal{V}$$

Then

$$0 \leqslant U^* - \left\langle \theta, \frac{\mu^{1,t} + \mu^{2,t}}{2} \right\rangle \to 0 \tag{2}$$

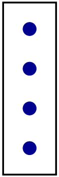
Implementation: Numerical Issues

$\theta_{c,1}$

$\theta_{c,2}$

$\vdots$

$\theta_{c,n}$



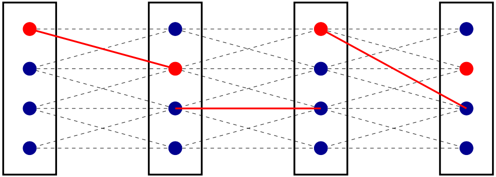
$\rightarrow$?

$\exp(\theta_{c,1})$

$\vdots$

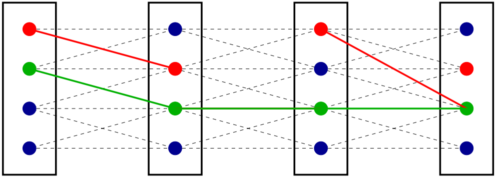
$\exp(\theta_{c,n})$

$$\theta^*_{c,i} = \theta_{c,i} - \max_{j=\overline{1,n}} \theta_{c,j}$$



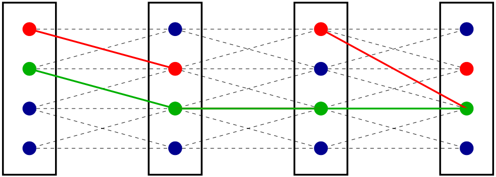
$$\exp\langle \theta^*/\rho, \phi(y)\rangle \xrightarrow[\rho\to 0]{} 0 \quad \forall y \in \mathcal{Y}^{\mathcal{V}}$$

Implementation: Numerical Issues



$$\theta^*_{c,y} = \theta_{c,y} - \theta_{c,y^*}$$
$$y^* = \arg\max_{y \in \mathcal{Y}^\mathcal{V}} \langle \theta, \phi(y) \rangle$$
$$\Rightarrow \ \exists c \in \mathcal{C}: \ \max_{y \in \mathcal{Y}} \theta^*_{c,y} > 0$$

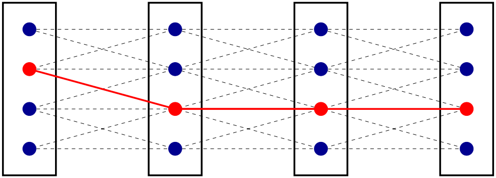
Implementation: Numerical Issues



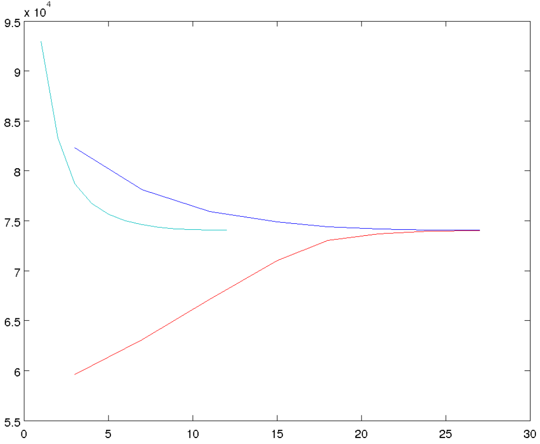
$$\theta^*_{c,y} = \theta_{c,y} - \theta_{c,y^*}$$
$$y^* = \arg\max_{y \in \mathcal{Y}^{\mathcal{V}}} \langle \theta, \phi(y) \rangle$$
$$\Rightarrow \; \exists c \in \mathcal{C}: \; \max_{y \in \mathcal{Y}} \theta^*_{c,y} > 0$$

Solution: an equivalent transformation $\theta \to \theta^*$:

$$\langle \theta, \phi(y) \rangle = \langle \theta^*, \phi(y) \rangle \;\; \forall y \in \mathcal{Y}^{\mathcal{V}}$$
$$y^* \in \operatorname{Arg}\max_{y \in \mathcal{Y}^{\mathcal{V}}} \langle \theta^*, \phi(y) \rangle \Leftrightarrow \forall c \in \mathcal{C}: \;\; \theta^*_{c,y^*} \in \operatorname{Arg}\max_{y \in \mathcal{Y}} \theta^*_{c,y}$$

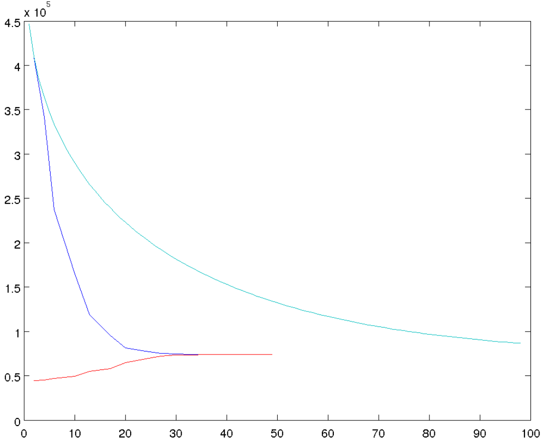


Random Sample: 5 Labels, an LP Tight Case. Close to a Trivial Problem



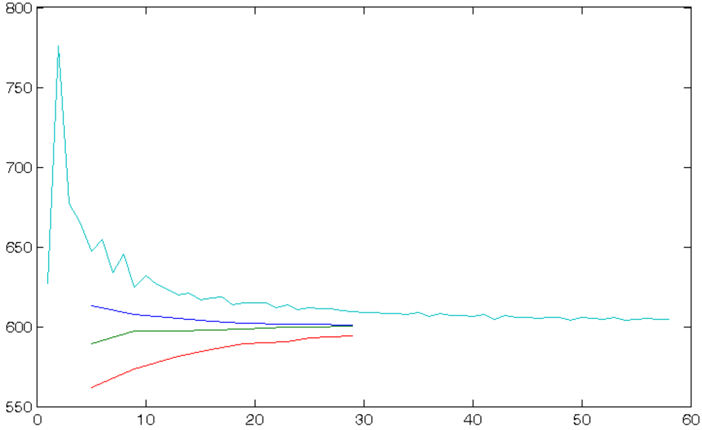
Checked on different image size from 20×20 to 200×200 and different distribution of weights between pairwise and unary factors.

Random Sample: 5 Labels, an LP Tight Case. Far from a Trivial Problem

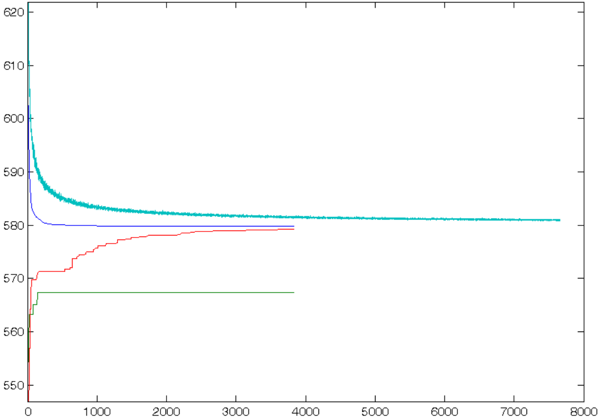


Checked on different image size from 20×20 to 200×200 and different distribution of weights between pairwise and unary factors.

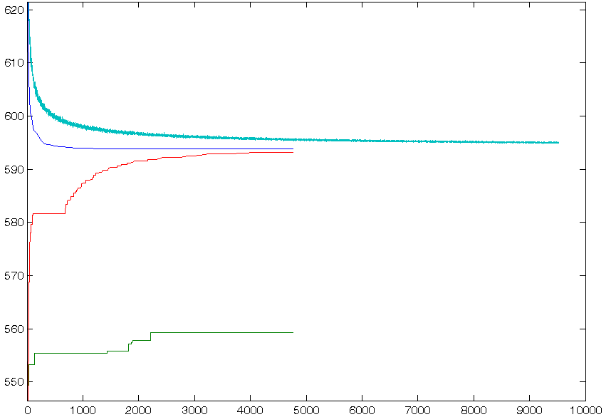
Random Sample: 5 Labels, Uniform Distribution with pw/u Weights 0.01



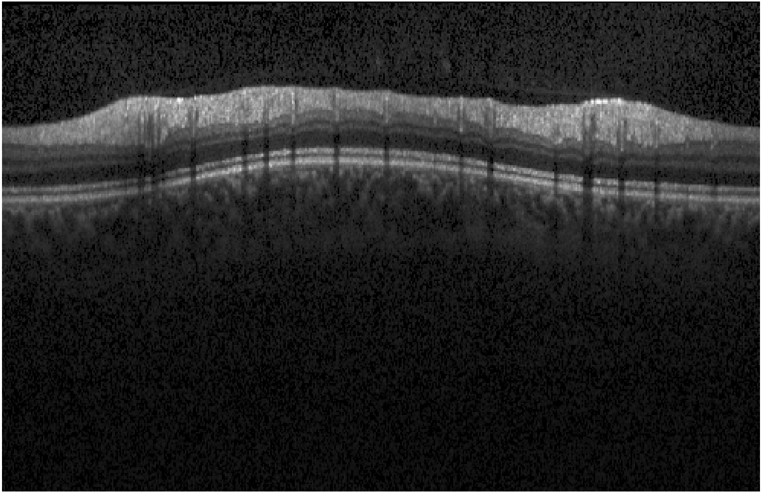
Random Sample: 5 Labels, Uniform Distribution with pw/u Weights 0.50



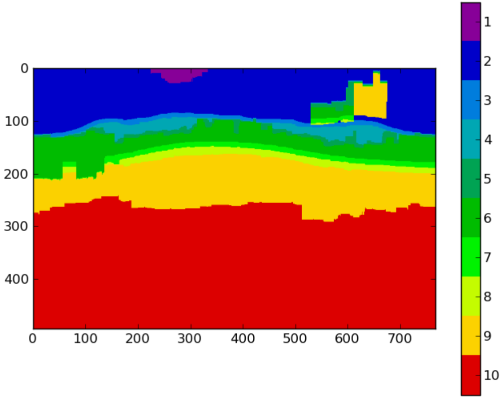
Random Sample: 5 Labels, Uniform Distribution with pw/u Weights 0.75



Input image



Precision 0.01% (227 oracle calls)



Summary

We proposed:

1. A smooth optimal first-order optimization method to solve the MAP inference problem.
2. A method for the lower bound calculation for non-LP-tight case.
3. Dynamic Lipschitz constant estimation to significantly speed-up the calculations.
4. A non-trivial implementation for (almost) any arbitrary small smoothing value.

Future work

1. Further speed-up.
2. Deeper insight to a way of changing the smoothing parameter in the course of the algorithm.
3. Improvement of convergence of the primal objective for a non-LP-tight cases.

Analysis of paper by Jojic, Gould, Koller: ICML 2010

+ Clever idea!
- Lipschitz constant is estimated for L_1-norm and algorithm (seems that) uses L_2-norm - error !
- Stopping criterion is not specified.
- Numerical issues are not covered: experiments only with decompositions to small subgraphs.

Optimization: Estimate Sequence Notion [Nesterov83]

1. $\{g^t(x)\}, \alpha^t \in (0,1),\ \sum_{t=0}^{\infty} = \infty, \quad t = \overline{1,\infty}$

$$g^{t+1}(x) \leqslant \alpha^t f(x) + (1-\alpha^t) g^t(x)$$

2. if for some $\{x^t\}$ $f(x^t) \leqslant g^{t*} \equiv \min_{x \in \mathbb{R}^n} g^t(x)$ then

$$f(x^t) - f^* \leqslant \lambda^t (g_0(x^t) - f^*) \to 0, \text{for } \lambda^t = \prod_{t=0}^{\infty} (1-\alpha^t) \to 0.$$

Optimization: Estimate Sequence Example [Nesterov83]

1. f-convex $\Rightarrow$ $f(x) \geqslant f(z) + \langle f'(z), x - z \rangle \Rightarrow$ for any sequence $\{z^t\}$

$$g^{t+1}(x) = \alpha^t (f(z) + \langle f'(z^t), x - z \rangle) + (1 - \alpha^t) g^t(x)$$

2. let $g_0(x) = g_0^* + \frac{\gamma_0}{2} ||x - u_0||^2$ then

$$g^t(x) = g^{t*} + \frac{\gamma^t}{2} ||x - u^t||^2$$

and for g^{t*}, γ^t, u^t there are closed form expressions.

3. We have to define sequences a^t, z^t and x^t such that $f(x^t) \leqslant g^{t*}$.
4.

$$f(x^t) - f^* \leqslant \lambda^t \left(f(x_0) - f^* + \frac{\gamma_0}{2} ||x_0 - x^*||^2 \right) \rightarrow 0$$

Wainwright Martin J., Jaakkola Tommi S., and Willsky Alan S.
Map estimation via agreement on trees: Message-passing and linear programming.
IEEE Trans. on Inf. Theory, 51(11), November 2005.

Nikos Komodakis, Nikos Paragios, and Georgios Tziritas.
MRF optimization via dual decomposition: Message-passing revisited.
In *ICCV*, 2007.

Schlesinger Michail and Giginiak Volodymyr.
Solution to structural recognition (max,+)-problems by their equivalent transformations.
Control Systems and Computers, (1-2), 2007.

Yurii Nesterov.
A method for solving a convex programming problem with convergence rate $1/k^2$.
Soviet Math. Dokl., 27(2):372–376, 1983.

Yurii Nesterov.
Introductory Lectures on Convex Optimization: A Basic Course.

Kluwer Academic Publishers, Boston/Dordrecht/London, 2004.

Yurii Nesterov.
Smooth minimization of non-smooth functions.
Math. Program., Ser. A(103):127–152, 2004.

Yurii Nesterov.
Gradient methods for minimizing composite objective function.
CORE Discussion Paper 2007/76, page 30, 2007.

Tomas Werner.
A linear programming approach to max-sum problem: A review.
IEEE Trans. on Pattern Recognition and Machine Intelligence (PAMI), 29(7), July 2007.

Tomas Werner.
Revisiting the decomposition approach to inference in exponential families and graphical models.
Technical report, Center for Machine Perception, Czech Technical University, 2009.