Обзор методов решения задачи MAP

А. В. Новиков[1]

[1]МГУ, ВМиК, каф. ММП

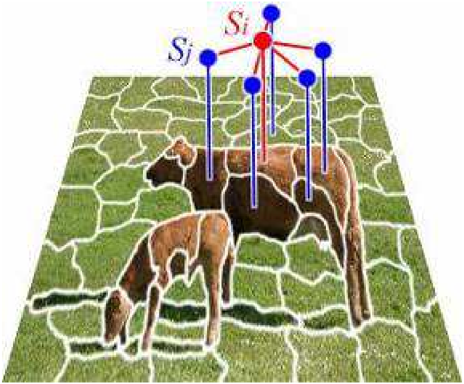
Спецсеминар «Байесовские методы машинного обучения»

Оглавление

Постановка задачи

$$P(x \mid z) \propto \prod_{a \in V} \psi(z_a \mid x_a) \prod_{(a,b) \in E} \psi_{ab}(x_a, x_b) \to \max_X$$

$$E(x) = \sum_{a \in V} \varphi_a(x_a) + \sum_{(a,b) \in E} \varphi_{ab}(x_a, x_b) \to \min_X \quad (1)$$

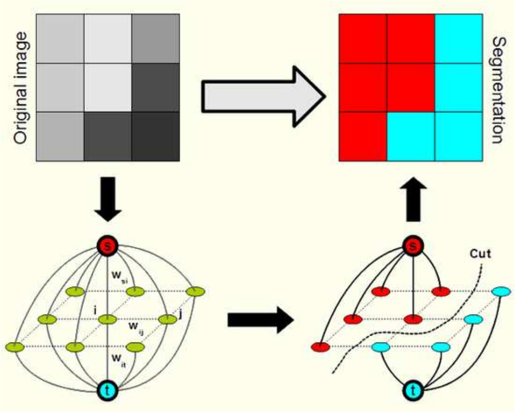


Обзор алгоритмов

- Alpha–Expansion
- Loopy belief propagation
- TRW–S [1]
- Субградиентный подъём [4]
- Bundle methods [5]
- L–BFGS

Разрез графов

В случае если задача бинарная ($K = 2$) и выполнено «условие субмодулярности» ($\varphi_{ab}(0,0) + \varphi_{ab}(1,1) \leq \varphi_{ab}(0,1) + \varphi_{ab}(1,0)$), то можно применять очень эффективные алгоритмы основанные на разрезе графов.



Двойственное разложение (пример)

Пусть у нас есть задача

$$f(x) + g(x) \to \min_x$$

причем задачи

$$f(x) \to \min_x$$
$$g(x) \to \min_x$$

мы умеем решать быстро. Тогда:

$$f(x_1) + g(x_2) \to \min_{x_1, x_2}$$
$$\textit{such that } x_1 = x_2$$

$$\max_\lambda \min_{x_1, x_2} f(x_1) + g(x_2) + \lambda(x_2 - x_1)$$

$$\max_\lambda \min_{x_1, x_2} f(x_1) - \lambda x_1 + g(x_2) + \lambda x_2$$

Overcomplete representation

Введем дополнительные переменные:

- $x_a \to \{y_{a,1} \dots y_{a,K}\}$: $y_{a,p} = 1 \Leftrightarrow x_a = p$
- $\{y_{ab,11}, y_{ab,12} \dots, y_{ab,KK}\}$: $y_{ab,pq} = 1 \Leftrightarrow x_a = p, x_b = q$
- $\theta_{a,p} = \varphi_a(p)$, $\theta_{ab,pq} = \varphi_{ab}(p,q)$

Тогда:

$$E(Y,\Theta) = \sum_{a \in V} \sum_{p=1}^{K} \theta_{a,p} y_{a,p} + \sum_{(a,b) \in E} \sum_{p,q=1,1}^{K} \theta_{ab,pq} y_{ab,pq} \to \min_{Y \in \mathcal{M}} \qquad (2)$$

$$\mathcal{M} = \Bigg\{ Y \mid y_{a,p}, y_{ab,pq} \in \{0,\ 1\}, \sum_{p=1}^{K} y_{a,p} = 1, \\ \sum_{p=1}^{K} y_{ab,pq} = y_{bq}, \sum_{q=1}^{K} y_{ab,pq} = y_{ap} \Bigg\}$$

LP–релаксация

$$\mathcal{R} = \left\{ Y \mid y_{a,p}, y_{ab,pq} \in [0,\ 1], \sum_{p=1}^{K} y_{a,p} = 1, \right.$$
$$\left. \sum_{p=1}^{K} y_{ab,pq} = y_{b,q}, \sum_{q=1}^{K} y_{ab,pq} = y_{a,p} \right\}$$

$$\min_{y \in \mathcal{M}} E(Y, \Theta) \geq \min_{Y \in \mathcal{R}} E(y, \Theta)$$

При этом если граф является деревом достигается равенство.

Двойственное разложение 2

Разобьем граф G на деревья $\{D^t\}_{t=1}^T$ так, чтобы каждая вершина и каждое ребро G входила ходя бы в одно дерево. n_a — число подграфов включающих вершину a.

$$\theta^t_{a,p} = \begin{cases} \frac{\theta_{a,p}}{n_a}, & a \in D^t \\ 0, & a \notin D^t \end{cases}$$

Аналогично, n_{ab} — число деревьев включающих ребро (a, b),

$$\theta^t_{ab,pq} = \begin{cases} \frac{\theta_{ab,pq}}{n_{ab}}, & (a,b) \in D^t \\ 0, & (a,b) \notin D^t \end{cases}$$

$$\Theta = \sum_{t=1}^{T} \Theta^t$$

$$E(Y, \Theta) = \sum_{t=1}^{T} E(Y, \Theta^t)$$

Двойственное разложение 3

Введем дополнительные переменные
$\Lambda = \{\Lambda^t\}_{t=1}^T = \{\{\lambda_{a,p}^t\}, \{\lambda_{ab,pq}^t\}\}_{t=1}^T \in \mathcal{L}$, где $\mathcal{L}$ задается ограничениями:

$$\sum_{t=1}^{T} \lambda_{a,p}^t = 0, \forall a, p$$

$$\sum_{t=1}^{T} \lambda_{ab,pq}^t = 0, \forall (a, b), p, q$$

Собирая всё вместе:

$$\min_{Y \in \mathcal{M}} E(Y \mid \Theta) \geq \min_{Y \in \mathcal{R}} E(Y \mid \Theta) =$$

$$= \min_{Y \in \mathcal{R}} E(Y \mid \Theta) + \sum_{t=1}^{T} \left[\sum_{a \in V} \sum_{p=1}^{K} \lambda_{a,p} y_{a,p} + \sum_{(a,b) \in E} \sum_{p,q=1}^{K} \lambda_{ab,pq} y_{ab,pq} \right] =$$

$$= \min_{Y \in \mathcal{R}} E(Y \mid \Theta + \Lambda) \geq \sum_{t=1}^{T} \min_{Y \in \mathcal{R}} E(Y \mid \Theta^t + \Lambda^t)$$

Двойственное разложение 4

В каждом слагаемом у нас LP–релаксация задачи (2) для дерева, а значит:

$$\sum_{t=1}^{T} \min_{Y \in \mathcal{R}} E(Y \mid \Theta^t + \Lambda^t) = \sum_{t=1}^{T} \min_{Y \in \mathcal{M}} E(Y \mid \Theta^t + \Lambda^t) = g(\Lambda) \qquad (3)$$

Заметим, что $\min_{Y \in \mathcal{M}} E(Y \mid \Theta^t + \Lambda^t)$ является минимум конечного (хотя и очень большого) числа линейных по Λ^t функцией, то–есть вогнутой функцией.

Двойственное разложение 5 (субградиенты)

Не сложно посчитать проекцию субградиента функции $g(\Lambda)$ на множество $\mathcal{L}$:

$$P_{\mathcal{L}}\left(\frac{\partial}{\partial \lambda^t_{a,p}} g(\Lambda)\right) = \widehat{y}^t_{a,p} - \frac{\sum_{\{t' \mid a \in D^{t'}\}} \widehat{y}^{t'}_{a,p}}{n_a} \tag{4}$$

, где $\widehat{y}^t_{ap} = arg\,\min_{Y \in \mathcal{M}} E(Y \mid \Theta^t + \Lambda^t)$, то есть решение внутренней задачи минимизации на дереве. Аналогично,

$$P_{\mathcal{L}}\left(\frac{\partial}{\partial \lambda^t_{ab,pq}} g(\Lambda)\right) = \widehat{y}^t_{ab,pq} - \frac{\sum_{\{t' \mid (a,b) \in D^{t'}\}} \widehat{y}^{t'}_{ab,pq}}{n_{ab}} \tag{5}$$

, где $\widehat{y}^t_{ab,pq} = arg\,\min_{Y \in \mathcal{M}} E(Y \mid \Theta^t + \Lambda^t)$.

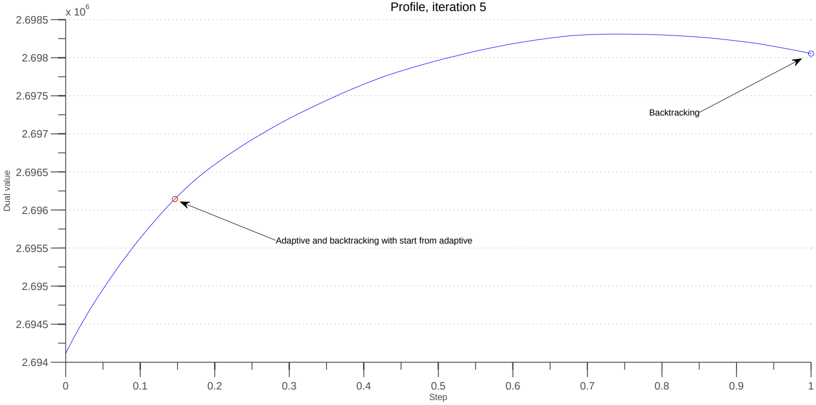
Субградиентный подъем

В этом случае мы просто идем в сторону субградиента и метод полностью определяет последовательность шагов. Мы опробовали:

- Константный шаг
- Адаптивный шаг
- Метод Флетчера
- Backtracking
- Точная оптимизация

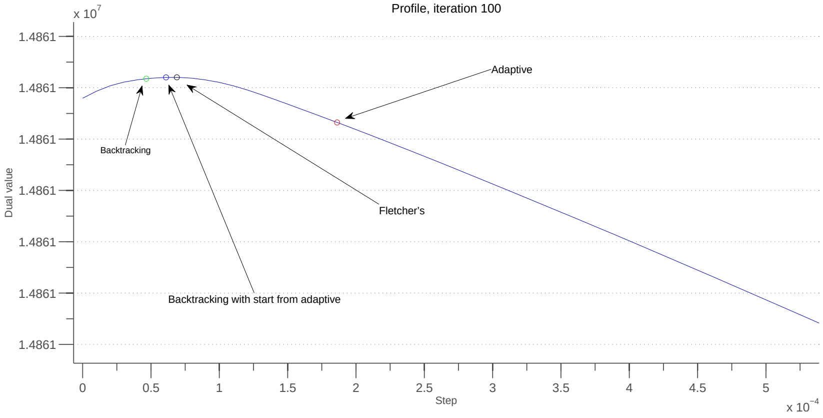
Субградиентный подъем

Профиль 1



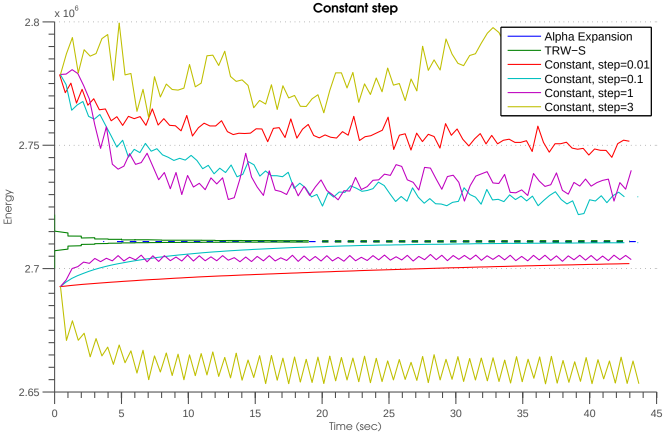
Субградиентный подъем

Профиль 2



Субградиентный подъем

Константный шаг



Субградиентный подъем

Адаптивный выбор шага

$$\alpha^k = \frac{Approx^k - Dual^k}{\|P_{\mathcal{L}}(g^k)\|^2} \tag{6}$$

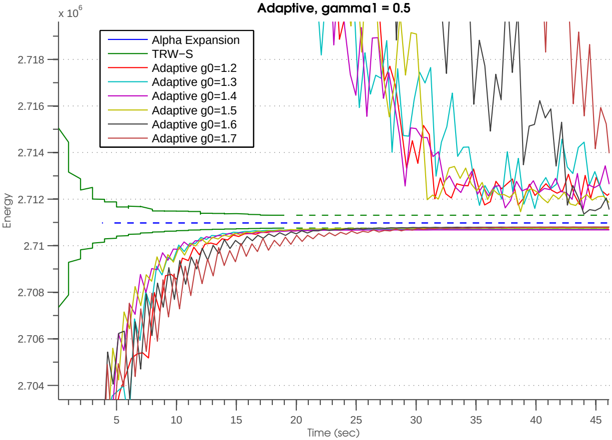
Где $Dual^k$ — текущее значение двойственной функции, $Approx^k$ — оценка оптимума двойственной функции.
$Approx^k = BestDual^k + \delta^k$,
где $BestDual^k$ — лучшее на данный момент значение двойственной функции,

$$\delta^{k+1} = \begin{cases} \gamma_0 \delta^k, & Dual^k > Dual^{k-1}, \\ max(\gamma_1 \delta^k, \epsilon) & Dual^k \leqslant Dual^{k-1}. \end{cases} \tag{7}$$

γ_0, γ_1, ϵ — параметры метода, выбранные нами эмпирически.
$\gamma_0 = 1.4$
$\gamma_1 = 0.5$
$\epsilon^k = \frac{1}{k}$

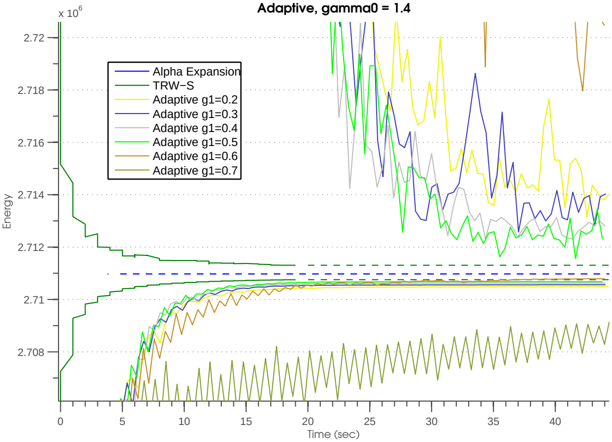
Субградиентный подъем

Адаптивный выбор шага (γ^0)



Субградиентный подъем

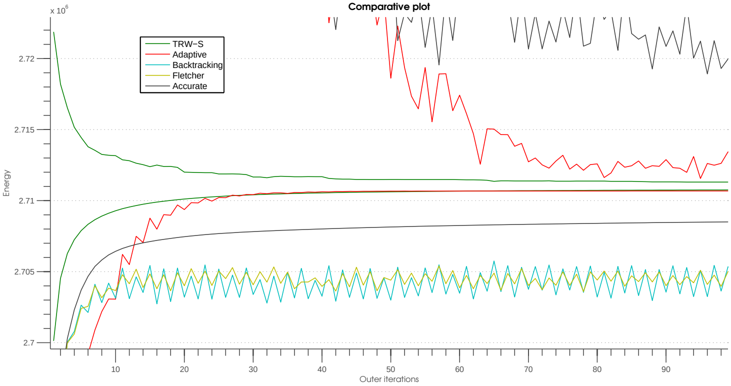
Адаптивный выбор шага (γ_1)



Субградиентный подъем

Сравнение подходов

Обратите внимание на то, что по оси X отложенны внешние итерации метода.



Bundle approach

Построим простую глобальную оценку сверху и будем оптимизировать её:

$$\hat{f}(\lambda) = \min_{(\lambda', f(\lambda'), g') \in \mathcal{B}} \{ f(\lambda') + < g', \lambda - \lambda' > \} \tag{8}$$

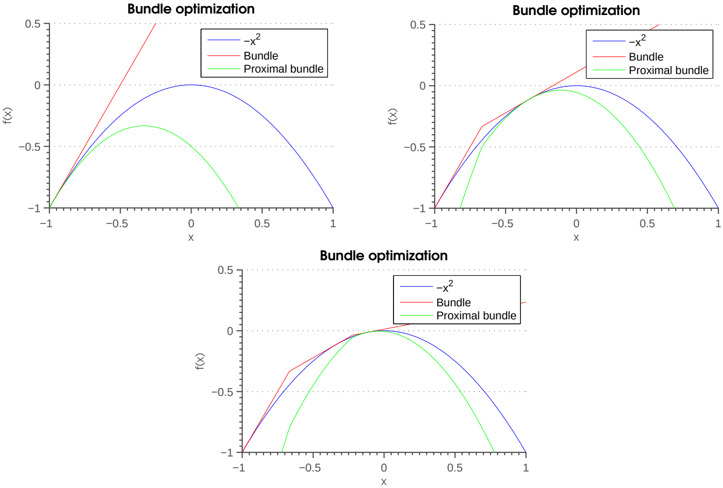
$$\lambda^{k+1} = arg \max_{\lambda} \{ \hat{f}(\lambda) - \frac{w^k}{2} \left\| \lambda - \overline{\lambda} \right\|_2^2 \} \tag{9}$$

Качество работы алгоритма зависит от управления размером бандла и выбора весов. Авторы метода предлагают следующую формулу пересчета весов [5]:

$$w^k = P_{[w_{min}, w_{max}]} \left(\left(\gamma \cdot \frac{\min_k Upper^k - \max_k Dual^k}{\| P_{\mathcal{L}}(g^k) \|} \right)^{-1} \right) \tag{10}$$

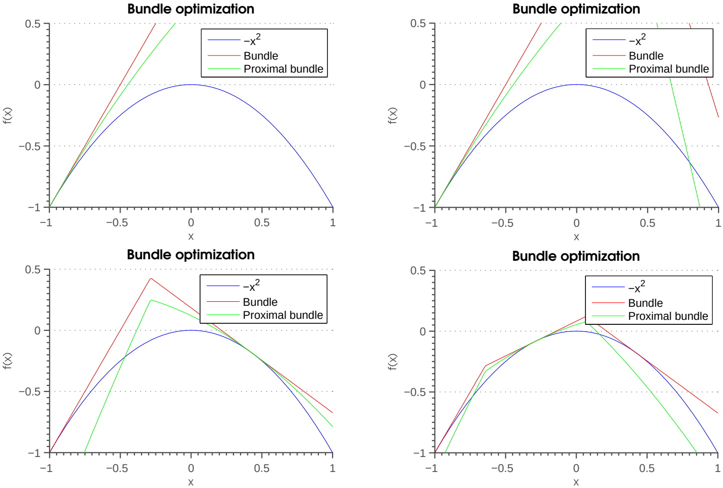
Bundle approach, пример 1

Рассмотрим вариант с весом равным довольно большой константе 3.



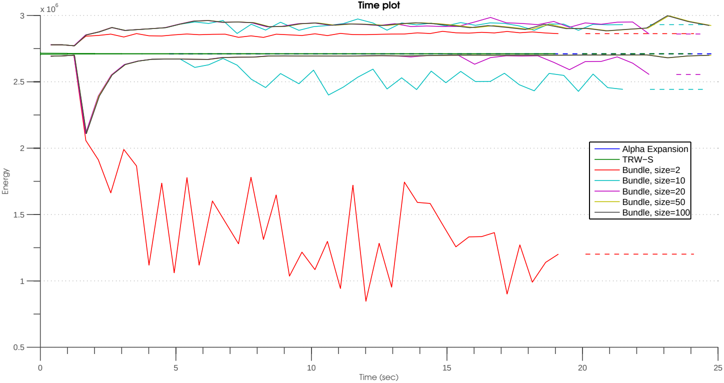
Bundle approach, пример 2

Теперь весом будет маленькая константа 0.7 и шаги увеличатся.



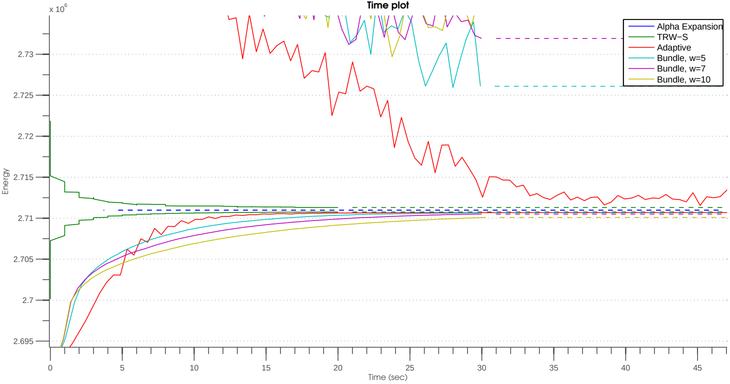
Bundle approach

Воспользуемся формулой от авторов и поварьируем размер бандла.
При этом w^k становится порядка 0.01.

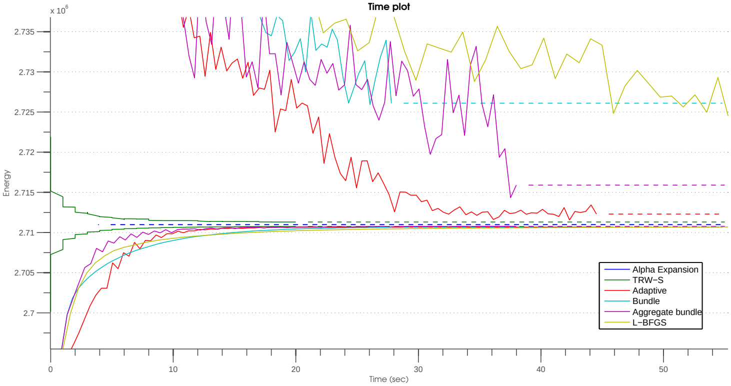


Bundle approach

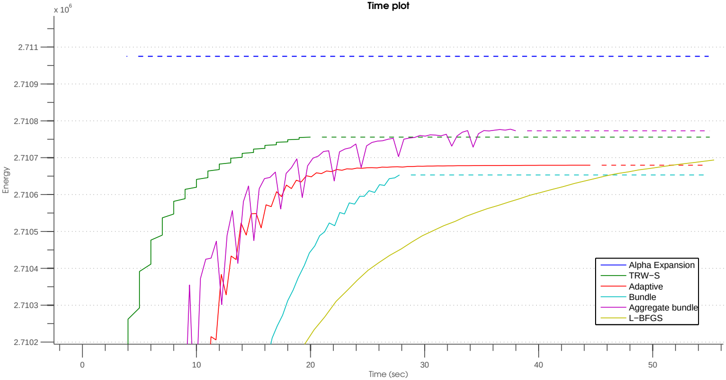
Теперь зафиксируем вес w^k огромной константой (размер бандла = 10).



Сравнение методов

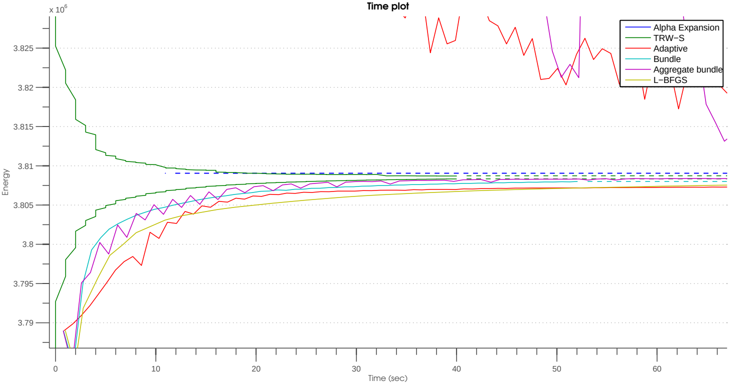


Сравнение методов



Сравнение методов

Для стерео пары venus:



Выводы, вклад

- Реализован фреймворк для сравнения алгоритмов вывода в MRF
- Лучше всего работают aggregate bundle методы, тема ещё не до конца закончена (проблемы с выбором шага)
- На удивленее неплохо работает метод адаптивного субградиента

- Kolmogorov V. Convergent Tree-Reweighted Message Passing for Energy Minimization // IEEE Trans. Pattern Anal. Mach. Intell., 2006. — С. 1568–1583.
- Kiwiel K. An aggregate subgradient method for nonsmooth convex minimization // Mathematical Programming, 1983, 27:320–341.
- Alahari K., Kohli P., Torr P. H. S. Dynamic Hybrid Algorithms for MAP Inference in Discrete MRFs // IEEE Trans. Pattern Anal. Mach. Intell., 2010. — С. 1846–1857.
- Komodakis N., Paragios N., Tziritas G. MRF energy minimization and beyond via dual decomposition // Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011. — С. 531–552.
- Kappes J. H., Bogdan Savchynskyy, Christoph Schnorr A Bundle Approach To Efficient MAP-Inference by Lagrangian Relaxation // Computer Vision and Pattern Recognition (CVPR), IEEE Conference 2012. — С. 1688–1695.