Lecture 5
★
Approximate (and exact) minimization of functionals in image analysis with graph cuts

Erik L. G. Wernersson
erikw@cb.uu.se

2012-05-03

Goals

- Derive an interesting functional for binary images
- Show how that one and similar functionals can be converted into graph cut problems.
- Provide some tools to approach a general functional of binary functions.
- Introduce the more general, and interesting, N-label problems.
- Give and informative, informal and interesting lecture!

Contents

- Background and repetition.
- An example leading to a functional that we would like to minimize.
- Minimization with graph cuts.
- Graph representability of problems.
- Many label problems.
- Some related topics.

Minimal s-t-cuts

Let $G = (V, E)$ be a directed graph with positive edge weights and two special nodes, $s, t \in V$. An s-t-cut is a partition of V into S, T such that $S \cup T = V$, $S \cap T = \emptyset$ and $s \in S$ and $t \in T$. The cut is associated with a cost

$$c(S, T) = \sum_{u \in S, v \in T, (u,v) \in E} c(u, v)$$

The minimal s-t-cut is the cut C that minimizes c. This can be solved by a max flow algorithm (different varieties).

Max flow
Ford-Fulkerson's algoritm

Require: $G(V, E)$, $s, t \in V$
  while $\exists$ path, p from s to t **do**
    $c(p) = \min c(u, v), \quad (u, v) \in p$
    for $\forall\, (u, v) \in p$ **do**
      $f(u, v) = f(u, v) + c(u, v)$
    end for
  end while
  return f

⋆

Maximum A Posteori Restoration of Images, leading to an interesting functional

"interesting functionals are often difficult to minimize"

How can we segment an image like this into two classes?

Derivation of a two label problem
Following (Greig, 1989)

Given an observation y find the most probable image $\hat{x} : N \times N \rightarrow \{0, 1\}$ by maximizing $p(x|y) \propto p(y|x)p(x)$. The a priori probability of an image $p(x)$ is modeled by a MRF.

$$p(y|x) = \prod_{i=1}^{n} f(y_i|x_i) = \prod_{i=1}^{n} f(y_i|1)^{x_i} f(y_i|0)^{1-x_i}$$

$$p(x) \propto \exp\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}\left\{x_i x_j + (1-x_i)(1-x_j)\right\}\right]$$

$$\beta_{ij} = \beta, \quad \text{if}, \quad x_i \in \mathcal{N}(x_j), \quad \text{else}, \quad 0$$

Derivation of a two label problem
Log likelihood

erikw@cb.uu.se

Taking the logarithm of p gives the *log likelihood*, L. Both p and L are maximized by the same $\hat{x}$.

$$L(x|y) = k_0 + \sum_{i=1}^{n} \lambda_i x_i + \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{m} \beta_{ij} \left\{ x_i x_j + (1 - x_i)(1 - x_j) \right\}$$

$$\lambda_i = \ln \frac{f(y_i|1)}{f(y_i|0)}$$

Derivation of a two label problem
Graph construction

Construct a graph with one node for each pixel and one s and t node such that

$$c_{si} = \lambda_i \quad \text{if} \quad \lambda_i > 0,$$

$$c_{it} = -\lambda_i \quad \text{if} \quad \lambda_i < 0,$$

$$c_{ij} = c_{ji} = \beta \quad \text{if} \quad (i, j) \in \mathcal{N}$$

A cut will give two partitions, $B = \{s\} \cup \{i : x_i = 1\}$ and $W = \{t\} \cup \{i : x_i = 0\}$ and the capacity of the cut on an image x is

$$C(x) = \sum_{k \in B} \sum_{l \in W} c_{kl}$$

Derivation of a two label problem
Equivalence

A paper and pen exercise shows that the cut can be expressed as

$$C(x) = \sum_{i=1}^{n} \left[x_i \max(0, -\lambda_i) + (1 - x_i) \max(0, \lambda_i)\right] \ldots$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} (x_i - x_j)^2$$

and further that C and L are equal up to a constant and a sign change:

$$-L(x|y) = C(x) + k_2$$

Maximizing $L(x, y)$ is equivalent to finding the minimum cut $C(x)$ in the network.

Example

Say we have an image which we know consists of piecewise flat objects but is corrupted with Gaussian iid $\in (0, \sigma)$ where object has intensity α and background γ, then,

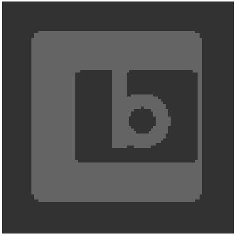
$$p(y_i|\text{obj}) \propto e^{-\frac{(y_i-\alpha)^2}{2\sigma^2}}$$

and

$$\lambda_i = \ln \frac{f(y_i|\text{obj})}{f(y_i|\text{bg})} = \ln e^{\frac{-(y_i-\alpha)^2}{2\sigma^2}} - \ln e^{-\frac{(y_i-\gamma)^2}{2\sigma^2}}$$

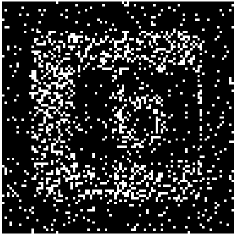
Example cont.

Reference



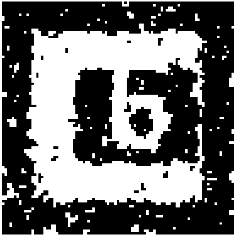
Observed

Thresholded

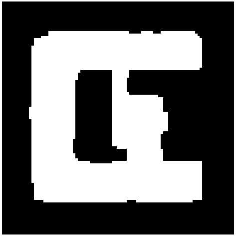


$\beta = \beta_0$

$\beta = 4\beta_0$



$\beta = 16\beta_0$



That functional was quite easy to minimize, right!?

$\star\star$

Functionals that can be represented by graphs and globally minimized

"researchers sometimes use heurestic methods for optimization, even in situations where the global minimum can be computed with graph cuts"

Functionals and minimization problems

Many functionals can be written in the following form

$$E(f) = E_{\text{individual}}(f) + E_{\text{interaction}}(f),$$

where $f \in \{0, 1\}$.

$$E_{\text{individual}} = \sum_{p \in P} D_p(f_p)$$

$$E_{\text{interaction}}(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q)$$

What functionals can be minimized?
(Kolmogorov 2004)

What kind of functionals on images can be formulated and solved with graphs and graphs cuts? Two classes have been studied and can be minimized under certain restrictions:

$\mathcal{F}^2$

$$E(x_1, ..., x_n) = \sum_i E^i(x_i) + \sum_{i<j} E^{i,j}(x_i, x_j)$$

$\mathcal{F}^3$

$$E(x_1, ..., x_n) = \sum_i E^i(x_i) + \sum_{i<j} E^{i,j}(x_i, x_j) + \sum_{i<j<k} E^{i,j,k}(x_i, x_j, x_k)$$

$\mathcal{F}^2$

A functional $F(x)$ is *graph representable* if there exist a graph such that the value of $F(x)$ is equal to the capacity of the cut $C = (S, T)$ plus a constant.

If a functional is graph representable, the minimal s-t cuts can be used to find the $\hat{x}$ that minimises F.

A function $F \in \mathcal{F}^2$ is graph representable iff it is *regular* i.e. if

$$E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0), \quad i < j$$

The sum of two graph representable functions is graph representable.

A constructive proof
(Kolmogorov, 2004)

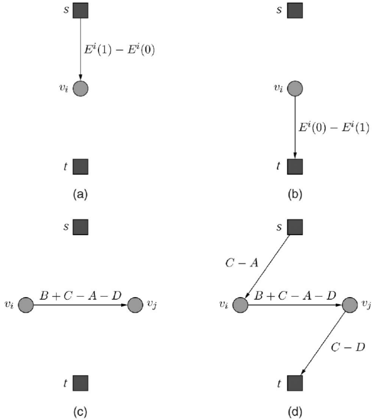


Fig. 2. Graphs that represent some functions in $\mathcal{F}^2$. (a) Graph for E^i, where $E^i(0) > E^i(1)$. (b) Graph for E^i, where $E^i(0) \not> E^i(0)$. (c) Third edge for $E^{i,j}$. (d) Complete graph for $E^{i,j}$ if $C > A$ and $C > D$.

TABLE 1

$$E^{i,j} = \begin{array}{|c|c|}\hline E^{i,j}(0,0) & E^{i,j}(0,1) \\ \hline E^{i,j}(1,0) & E^{i,j}(1,1) \\ \hline \end{array} = \begin{array}{|c|c|}\hline A & B \\ \hline C & D \\ \hline \end{array}$$

TABLE 2

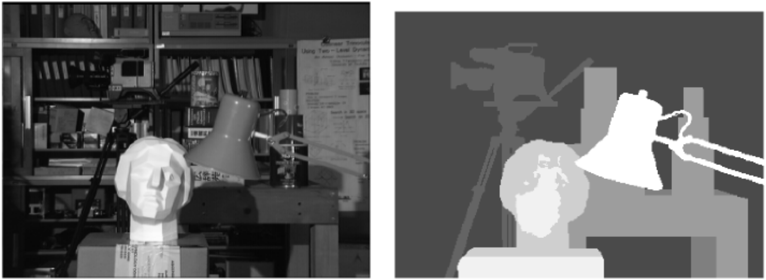
$$\begin{array}{|c|c|}\hline A & B \\ \hline C & D \\ \hline \end{array} = A + \begin{array}{|c|c|}\hline 0 & 0 \\ \hline C-A & C-A \\ \hline \end{array} + \begin{array}{|c|c|}\hline 0 & D-C \\ \hline 0 & D-C \\ \hline \end{array} + \begin{array}{|c|c|}\hline 0 & B+C-A-D \\ \hline 0 & 0 \\ \hline \end{array}$$

⋆ ⋆ ⋆

More on functionals and minimization, when there are more than two classes

"an approximate solution can be better than no solution"

Example
(Boykov, 2001)



Left stereo image and depth ground truth

The N Label Problem
with pairwise interaction

$$E(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p)$$

$$f \in \{0, 1, ..., N\}$$

V, the interaction penalty, is called a metric on the space of labels $\mathcal{L}$ if it satisfies 1–3 or semimetric if it satisfies 2–3:

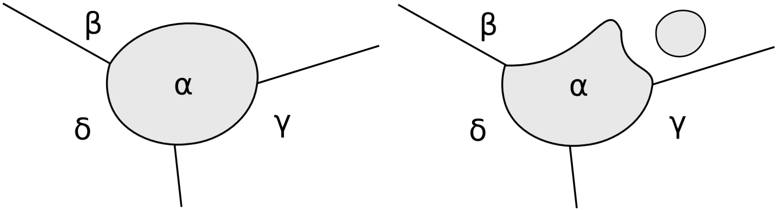
$$V(\alpha, \beta) = 0 \quad \Leftrightarrow \quad \alpha = \beta \tag{1}$$

$$V(\alpha, \beta) \quad = \quad V(\beta, \alpha) \geq 0 \tag{2}$$

$$V(\alpha, \beta) \quad \leq \quad V(\alpha, \gamma) + V(\gamma, \beta) \tag{3}$$

- The N Label problem is NP hard.
- There are two graph based methods that finds local minimas, but usually not the global minimas.
- These algorithms use “big” moves and produce good solution compared to methods that use “standard” moves”
- One of them has bounds for the maximum distance from the optimum.
- They are both based on ordinary s-t cuts.

α-β swap moves
For semi metric V



$\alpha - \beta$-swap

Definition

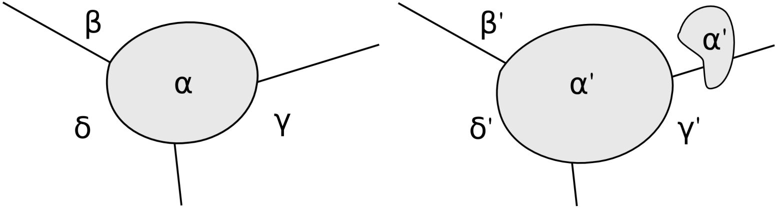
$$P_l = P'_l, l \neq \alpha, \beta$$

- only the partitions P_α and P_β change.

The swap algorithm
(Boykov, 2001)

1: Start with an arbitrary labelling, f
2: Set success $:= 0$
3: **for** Each pair of labels, $\{\alpha, \beta\} \subset \mathcal{L}$ **do**
4: Find $\hat{f} = \arg\min E(f')$ among f'
5: within one $\alpha - \beta$ swap of f.
6: **if** $E(\hat{f}) < E(f)$ **then**
7: $f := \hat{f}$, and success $:= 1$.
8: **end if**
9: **end for**
10: If success $= 1$ goto 2.
11: **return** f

α expansion moves
For metric V



α-expansion

Definition

$P_\alpha \subset P'_\alpha$ and $P'_l \subset P_l$

- Pixels are allowed to move from any partition to P_α.

The expansion algorithm
(Boykov, 2001)

1: Start with an arbitrary labelling, f
2: Set success $:= 0$
3: **for** Each label, $\alpha \in \mathcal{L}$ **do**
4: Find $\hat{f} = \arg\min E(f')$ among f'
5: within one α expansion move from f.
6: **if** $E(\hat{f}) < E(f)$ **then**
7: $f := \hat{f}$, and success $:= 1$.
8: **end if**
9: **end for**
10: If success $= 1$ goto 2.
11: **return** f

- The two algorithms have in common that only two labels at a time is allowed to change.
- Then the N-label problem is reduced to a two-label problem which there are exact solutions for.
- This does not guarantee that a global minima is found.
- How is convergence guaranteed?
- Can we say anything about the quality of the solutions?

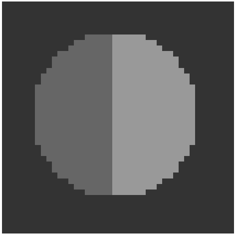
Optimality of the Expansion Move Algorithm

If $\hat{f}$ is the minimum found with the α expansion algorithm, and f^* is the global minimum then

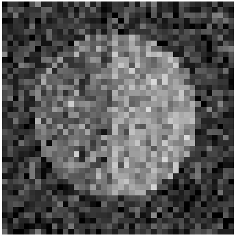
$$E(\hat{f}) \leq 2cE(f^*)$$

A Graphical Example

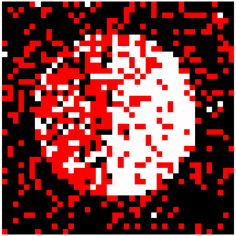
input



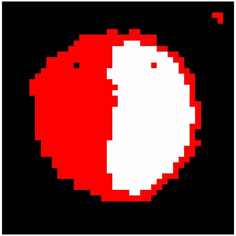
w. noise



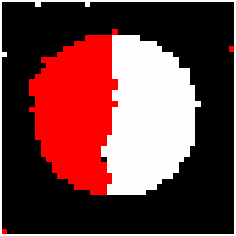
thresholded



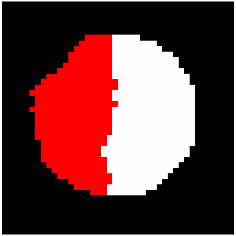
lp+thresholdning



α expansion



trunc. α expansion



Graph construction

- The α expansion and the α-β swap algorithms requires that intermediate graphs are constructed. This is slightly tricky, especially for α-β where new nodes are introduced at the boundaries.
- Code for multi label optimization (α expansion) can be downloaded from `http://vision.csd.uwo.ca/code/`.

⋆ ⋆ ⋆⋆

Related topics

Interaction terms

The Potts model:

$$V(\alpha, \beta) = K \cdot \delta(\alpha - \beta)$$

The truncated quadratics:

$$V(\alpha, \beta) = min(K, (\alpha - \beta)^2)$$

The truncated absolute distance:

$$V(\alpha, \beta) = min(K, |\alpha - \beta|)$$

Standard moves

To change one pixel at a time is called to make *standard moves*. Algoritms are based on them are bound to get stuck at really local minimas if energy descent is required.

- Iterated conditional modes (ICM) might be the simplest optimization method. It will operate on all problems that we have seen in the lecture.
- Simulated Annealing is also a candidate. In theory it can find the global minimum but in practice there is no example on when it is better than these graph cut methods.

What you can't optimize

Global properties can't be optimized with these techniques, like:

- Euler characteristics
- Circularity
- Number of objects
- ...

Summary

1. Many functionals can be minimized exactly with s-t cuts.
2. Even the potts model is NP-hard
3. Some NP hard problems is approximately solved using this cheap (low polynomial order) optimization routine iteratively. The solutions have bounds.
4. Any minimization based on graph cuts relies on intermediate binary construction.
5. Useful techniques, get to know them!

References

- **On The Statistical Analysis of Dirty Pictures**, J. Besag, *Journal of the Royal Statistical Society. Series B*, 48(3), 1986
- **Exact Minimim A Posteori Estimation for Binary Images**, D. M. Greig, B. T. Porteous and A. H. Seheult, *J. of the Royal Statistical Society. Series B.*, 51(2), 1989
- **An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision**, Y. Boykov and V. Kolmogorov, *Pattern Analysis and Machine Intelligence* 26(9), 2004
- **What Energy Functions can be Minimized via Graph Cuts?**, V. Kolmogorov and R. Zabih, *Pattern Analysis and Machine Intelligence* 26(2), 2004
- **Fast Approximate Energy Minimization via Graph Cuts**, Y. Boykov, O. Veksler, R. Zabih, *Pattern Analysis and Machine Intelligence* 23(11), 2001