Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M2. Optimization and inference techniques for Computer Vision**
nal exam
Date: 4rth December, 2014
Teachers: Coloma Ballester, Juan Fco Garamendi, Oriol Ramos, Joan Serra **Time: 2h30min**

- Books, lecture notes, calculators, phones, etc. are not allowed.
- All sheets of paper should have your name.
- Answer each problem in a separate sheet of paper.
- All results should be demonstrated or justified.

Problem 1 *1.5 Points*

Let

$$\begin{aligned} J:\ \mathcal{V} &\to \mathbb{R}, \\ u &\to J(u) = \int_\Omega \mathcal{F}\left(x, u(x), \nabla u(x)\right) dx \end{aligned}$$

be a convex energy functional over functions u, where

- $\mathcal{V}$ is a suitable space of functions.
- $\Omega \in \mathbb{R}^d$ is a bounded open domain of the d dimensional space $\mathbb{R}^d$.
- $u \in \mathcal{V}$, $u : \Omega \to \mathbb{R}$ is a scalar function defined on Ω.
- ∇ is the gradient operator and $x \in \Omega$ is the spatial variable.

(a) (0.15 points) Say in few words which is the fundamental problem in calculus of variations.

The fundamental problem of the calculus of variations is to find the extremum (maximum or minimum) of the functional $J(u)$.

(b) (0.15 points) Write the definition of the Gâteaux derivative of $J(u)$ (the directional derivative of J at u in direction h)

$$\left.\frac{dJ}{du}\right|_h = \lim_{\alpha \to 0} \frac{J(u + \alpha h) - J(u)}{\alpha}$$

(c) (0.15 points) Let

$$\frac{dJ}{du}$$

be the derivative of J at u. Which is the necessary condition for extremality of J?

$$\frac{dJ}{du} = 0$$

(d) (1.05 points) Apply the definition of Gâteaux derivative for finding the minimum at u of the following energy

$$J(u) = \frac{1}{2\lambda}\int_\Omega (u-f)^2\, dx$$

where $f(x) \in \mathcal{V}, f : \Omega \to \mathbb{R}$ is a fixed function.

First we use the Gâteaux derivative for obtaining the derivative of J at u.

$$\left.\frac{d\,J}{d\,u}\right|_h = \lim_{\alpha\to 0}\frac{1}{\alpha}\left(J(u+\alpha h) - J(u)\right)$$

That for the particular case of the given J

$$\left.\frac{d\,J}{d\,u}\right|_h = \lim_{\alpha\to 0}\frac{1}{\alpha}\left[\int_\Omega \frac{1}{2\lambda}((u+\alpha h)-f)^2 dx - \int_\Omega + \frac{1}{2\lambda}(u-f)^2 dx\right]$$

$$= \lim_{\alpha\to 0}\frac{1}{\alpha}\left[\frac{1}{2\lambda}\int_\Omega \left(((u-f)-\alpha h)^2 - (u-f)^2\right) dx\right]$$

$$= \lim_{\alpha\to 0}\frac{1}{\cancel{\alpha}}\left[\frac{1}{2\lambda}\int_\Omega \left(\cancel{(u-f)^2} - 2\cancel{\alpha}(u-f)h + \alpha^{\cancel{2}}h^2 \cancel{-(u-f)^2}\right) dx\right]$$

And going to the limit, the derivative of J at u in the direction of h is

$$\left.\frac{d\,J}{d\,u}\right|_h = \frac{1}{\cancel{2}\lambda}\int_\Omega -\cancel{2}(u-f)h dx$$

Recall that the directional derivative in the direction h can be written as the proyection of the derivative in that direction, thus

$$\left.\frac{d\,J}{d\,u}\right|_h = \left\langle \frac{d\,J}{d\,u}, h\right\rangle_{\mathcal{V}} = \int_\Omega \frac{d\,J}{d\,u} h dx$$

where $< f, g >_{\mathcal{V}}$ is the scalar product defined on $\mathcal{V}$

so

$$\frac{d\,J}{d\,u} = -\frac{1}{\lambda}(u-f)$$

As J is convex, J will have and extremum (in partiular the minimum) if and only if

$$\frac{d\,J}{d\,u} = f - u = 0$$

so $u = f$

Problem 2 *0.5 Points*

Let $f : \Omega \to \mathbb{R}$ be a given noisy image, where Ω is a bounded open subset of $\mathbb{R}^2$ and $f \in L^\infty(\Omega)$. Consider the following two minimization problems

P1:

$$\underset{u\in W^{1,2}(\Omega)}{\arg\min}\left\{\int_\Omega |\nabla u|^2 dx + \frac{1}{2\lambda_1}\int_\Omega |u-f|^2 dx\right\}$$

Where

- $W^{1,2}(\Omega) = \{u \in L^2(\Omega);\ \nabla u \in L^2(\Omega)^2\}$.
- $\lambda_1 \in \mathbb{R}^+$ is a given parameter.

P2:

$$\arg\min_{u\in BV(\Omega)} \left\{ \int_\Omega |Du| + \frac{1}{2\lambda_2}\int_\Omega |u-f|^2 dx \right\}$$

Where

- Du is the distributional gradient of u.
- $\int_\Omega |Du|$ is the total variation of u. If $u \in C^1(\Omega)$ (u smooth), $\int_\Omega |Du| = \int_\Omega |\nabla u| dx$.
- $BV(\Omega)$ is the space of bounded variation such as

$$BV(\Omega) = \left\{ u \in L^1(\Omega);\ \int_\Omega |Du| < \infty \right\}$$

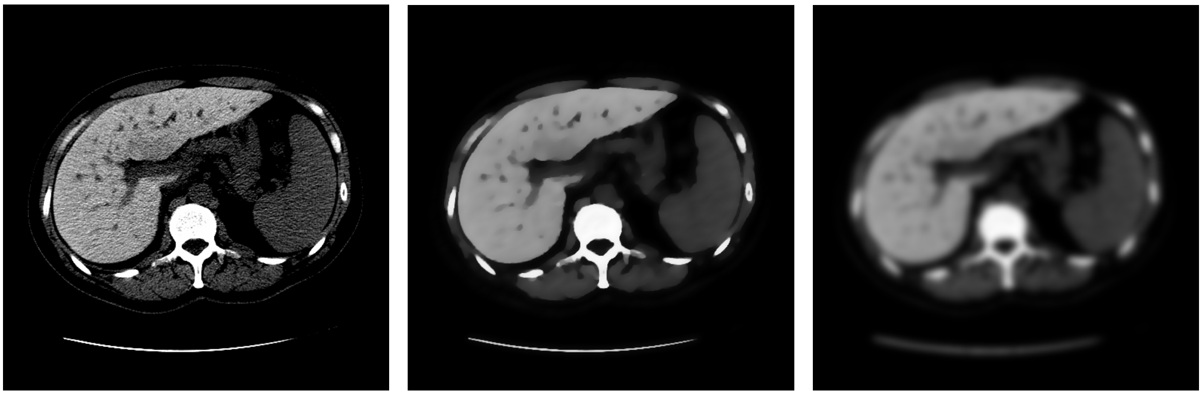
- $\lambda_2 \in \mathbb{R}^+$ is a given parameter.

Now, let $\lambda_1 = \lambda_2$ and consider the axial slice of a abdominal computed tomography shown at figure 1a as the noisy image f.

(a) (0.25 points) Say which image, figure 1b or figure 1c, corresponds to the solution of which problem P1 or P2. Justify your answer.

Hint: With abuse of notation, problem P2 can be written as $\arg\min_{u\in BV(\Omega)} \left\{ \int_\Omega |\nabla u| dx + \frac{1}{2\lambda_2}\int_\Omega |u-f|^2 dx \right\}$

Figure 1c is the solution for P1, and figure 1b is the solution for P2.
There are two possible answers for the justification
1.- The L^p norm with $p = 2$ of the gradient allows to remove the noise but penalizes too much the gradients corresponding to edges and the norm of the gradient in 1c are much more lower than in 1b.
2.- The solutions of problem P1 belongs to the space $W^{1,2}(\Omega)$, so they can not be solutions with discontinuities, as figure 1b presents.



(a) Given noisy image f | (b) Solution | (c) Solution

Figure 1: Figures for Problem 2

(b) (0.25 points) Which numerical and theoretical issue has the problem P2? Write at least one possible solution to overcome the issue.

The regularization term is not differentiable at origin, so it can not be formally differentiated, and the Euler-Lagrange can not be be formally written. One possible solution is regularize the total variation term as

$$TV_{\epsilon}(u) = \int_{\Omega} \sqrt{|\nabla u|^2 + \epsilon^2} dx$$

where $0 < \epsilon << 1$ is a given constant. Notice that when $\epsilon \to 0$, the written functional tends to Total Variation.
Other possible solution is to use the dual definition of total variation:

$$TV(u) := \sup_{\bar{p} \in P} \left\{ - \int_{\Omega} u \, \text{div}\bar{p} \, dx \right\}$$

Where
$\bar{p} := (p_1, p_2) : \Omega \to \mathbb{R}^2$ is a differentiable vector field with compact support ($\bar{p} = 0$ at $\partial\Omega$)
div$\bar{p}$ is the divergence operator such that div $\bar{p} := \sum_{j=1}^{D} \partial_j p_i$
$P := \{\bar{p} \in C_c^1(\Omega; \mathbb{R}^2) \ : \ |\bar{p}|_{L^\infty(\Omega)} \leq 1\}$ where

$$|\bar{p}| = \sup_{x \in \Omega} \sqrt{\sum_{i=1}^{2} p_i^2}$$

Problem 3 *0.5 Points*

For binary segmentation, the model problem can be defined as follows: Let $\omega \subset \Omega$ be an open, positive measured sub-region of the original domain (eventually not connected). If the curve Γ represents the boundary of such a segmentation ω then, in the level set formulation, the (free) moving boundary Γ is the zero level set of a Lipschitz function $\phi : \Omega \to \mathbb{R}$, that is:

$$\Gamma = \{(x, y) \in \Omega : \phi(x, y) = 0\}$$

where

$$\omega = \{(x, y) \in \Omega : \phi(x, y) > 0\}$$

and

$$\Omega \backslash \overline{\omega} = \{(x, y) \in \Omega : \phi(x, y) < 0\}$$

The level set function ϕ can be characterized as a minimum of the following energy functional,

$$J(\bar{c}, \phi) = \int_{\Omega} |DH(\phi)| + \lambda \int_{\Omega} (f - c_1)^2 H(\phi) \ + (f - c_2)^2 (1 - H(\phi)) dx$$

where $DH(\phi)$ is the distributional gradient of Heaviside function $H(\phi)$, f is a bounded function representing the image (the data), $\bar{c} \in \mathbb{R}^2$, $\bar{c} = (c_1, c_2)$ and $\lambda \in \mathbb{R}^+$ is a given parameter. The function $H(x)$ represents the Heaviside function, i.e.: $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ otherwise, and it allows to express the length of Γ by

$$|\Gamma| = \text{Length}(\phi = 0) = \int_{\Omega} |DH(\phi)|$$

where the term $\int_{\Omega} |DH(\phi)|$ denotes, properly, the total variation of the discontinuous function $H(\phi)$ in Ω.

(a) (0.25 points) We want to segment the image at figure 2a. For doing this, we minimize $J(\bar{c}, \phi)$ with two differents values of λ, named it λ_1 and λ_2, obtaing the segmentations at figure 2b and 2c. Consider $\lambda_1 >> \lambda_2$. Which λ corresponds with which result? Justify your answer.

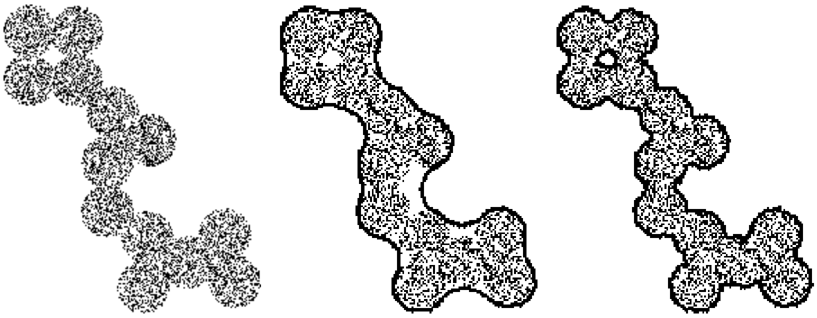
λ_1 corresponds to figure 2c and λ_2 corresponds to figure 2b. The reason is that for higher values of λ, the curve Γ, represented by the zero level set of ϕ, fits better to the data, or what is the same, lower values of λ allows smoother curves (or curves with lower lenght)

(b) (0.25 points) Why c_1 and c_2 are the mean values of the inside and outside regions?

Because when the energy functional is minimized w.r.t. c_1 and c_2, the necessary condition that is derivative of the energy functional w.r.t. c_1 and c_2 is zero give us

$$c_1 = \frac{\int_\Omega f H(\phi) dx}{\int_\Omega H(\phi) dx} \quad c_2 = \frac{\int_\Omega f\,(1 - H(\phi))\,dx}{\int_\Omega 1 - H(\phi) dx}$$

That are the the mean values.



(a) Image f to segment (b) Solution. $\phi = 0$ is drawn over f (c) Solution. $\phi = 0$ is drawn over f

Figure 2: Figures for Problem 3.

Problem 4 *2.5 Points*

A consultancy hires us to solve one of its problems. After some discussions, we have arrived to the following definitions of binary random variables (1=true, 0=false):

- x_1: A businessman corrupts a political leader,

$$p(x_1 = 1) = 0.30$$

- x_2: A political leader wins lottery,

$$p(x_2 = 1) = 0.01$$

- x_3: A businessman is in jail,

x_1	$p(x_3 = 1 \mid x_1)$
0	0.05
1	0.3

- x_4: A political leader is rich

x_1	x_2	$p(x_4 = 1 \mid x_1, x_2)$
0	0	0.1
0	1	0.2
1	0	0.6
1	1	0.9

- x_5: A political leader invests in a nightclub

x_2	$p(x_5 = 1 \mid x_2)$
0	0.01
1	0.1

- x_6: A political leader has a Swiss bank account,

x_4	$p(x_6 = 1 \mid x_4)$
0	0.05
1	0.1

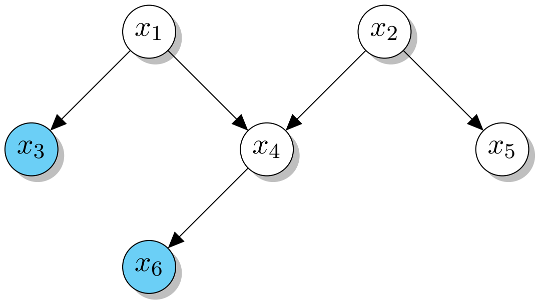
Moreover, we have the following observations:

- a businessman is in jail and

- a political leader has a Swiss bank account

a) Draw the Bayes Net of this problem and write the associated joint distribution (0.5 point).

Solution: The Bayes net is

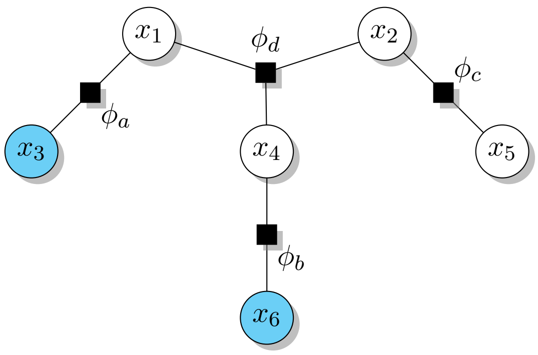


and the associated joint distribution:

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_3|x_1)p(x_6|x_4)p(x_4|x_1, x_2)p(x_5|x_2)p(x_1)p(x_2) \tag{1}$$

b) Convert the above Bayes net to a factor graph and define the corresponding factors, given the joint distribution defined in a) (0.5 point).

Solution: To convert the above Bayes net to a factor graph we have to convert each conditional probability $p(x_n|\text{pa}_n)$ to a factor function by *marrying* all the parents. This is done by defining a clique between a random variable and its parents, i.e.:



a possible factorization is the following:

- $\phi_a(x_1, x_3) = p(x_3|x_1)p(x_1)$
- $\phi_b(x_4, x_6) = p(x_6|x_4)$
- $\phi_c(x_2, x_5) = p(x_5|x_2)p(x_2)$
- $\phi_d(x_1, x_2, x_4) = p(x_4|x_1, x_2)$

c) Which is the complexity of the Belief Propagation (BP) algorithm when it is applied to any chain? and for the factor graph defined in b)? (0.5 point)

Solution: The complexity order of the BP algorithm for any PGM is $O(NK^c)$, where N is the number of random variables, K is the number of states and c is the size of the highest clique. For the factor graph defined in b, since we have 6 random variables, 2 states and the highest clique is 3, the complexity becomes: $O(6 \cdot 2^3)$

d) Estimate the probability that a political leader won a lottery given the above observations (1 point).

Solution: To estimate this probability we consider the variable x_2 as the *root* variable and we start by sending messages from *leaves* variables to factors and from factors to variables, until x_2 has received all the messages. Then, the probability is estimated by multiplying and normalizing all these messages.

Step by step, we proceed as follows. We compute the actual values of factor functions, given the probability tables above, and we initialize the messages depending on whether they were observed or not:

- $\phi_a(x_1, x_3) = p(x_3|x_1)p(x_1) = \begin{pmatrix} 0.95 & 0.05 \\ 0.70 & 0.30 \end{pmatrix} \circ \begin{pmatrix} 0.70 & 0.70 \\ 0.30 & 0.30 \end{pmatrix} = \begin{pmatrix} 0.665 & 0.035 \\ 0.21 & 0.09 \end{pmatrix}$
- $\phi_b(x_4, x_6) = p(x_6|x_4) = \begin{pmatrix} 0.95 & 0.05 \\ 0.90 & 0.10 \end{pmatrix}$
- $\phi_c(x_2, x_5) = p(x_5|x_2)p(x_2) = \begin{pmatrix} 0.99 & 0.01 \\ 0.90 & 0.10 \end{pmatrix} \circ \begin{pmatrix} 0.99 & 0.99 \\ 0.01 & 0.01 \end{pmatrix} = \begin{pmatrix} 0.98 & 0.01 \\ 0.01 & 0.00 \end{pmatrix}$
- $\phi_d(x_1, x_2, x_4) = p(x_4|x_1, x_2)$ is a "3 dimensional" array with the probabilities showed in the table above:

$$\phi_d(x_1, 0, x_4) = \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix} \text{ and } \phi_d(x_1, 1, x_4) = \begin{pmatrix} 0.80 & 0.20 \\ 0.10 & 0.90 \end{pmatrix}$$

The initial messages from the "leaves" variables are:

- $m_{3\to a} = m_{6\to b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
- $m_{5\to c} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Now we send messages from factors to variables:

$$m_{1\leftarrow a}(x_1) = \phi_a(x_1, x_3) m_{3\to a}(x_3) = \begin{pmatrix} 0.665 & 0.035 \\ 0.21 & 0.09 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.035 \\ 0.09 \end{pmatrix}$$
$$m_{4\leftarrow b}(x_4) = \phi_b(x_4, x_6) m_{6\to b}(x_6) = \begin{pmatrix} 0.95 & 0.05 \\ 0.90 & 0.10 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.10 \end{pmatrix}$$
$$m_{2\leftarrow c}(x_2) = \phi_c(x_2, x_5) m_{5\to c}(x_5) = \begin{pmatrix} 0.98 & 0.01 \\ 0.01 & 0.00 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.99 \\ 0.01 \end{pmatrix}$$

The next step is to send messages from variables to factors:

$$m_{1\to d}(x_1) = m_{1\leftarrow a}(x_1) = \begin{pmatrix} 0.035 \\ 0.09 \end{pmatrix}$$
$$m_{4\to d}(x_4) = m_{4\leftarrow b}(x_4) = \begin{pmatrix} 0.05 \\ 0.10 \end{pmatrix}$$

And we send the last message from factor d to variable x_2. Since d is a 3 order factor, we compute the messages for each value of x_2:

$$m_{2\leftarrow d}(0) = m_{1\to d}(x_1)^t \phi_d(x_1, 0, x_4) m_{4\to d}(x_4) = \begin{pmatrix} 0.035 & 0.09 \end{pmatrix} \begin{pmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{pmatrix} \begin{pmatrix} 0.05 \\ 0.10 \end{pmatrix} = 0.0091$$
$$m_{2\leftarrow d}(1) = m_{1\to d}(x_1)^t \phi_d(x_1, 1, x_4) m_{4\to d}(x_4) = \begin{pmatrix} 0.035 & 0.09 \end{pmatrix} \begin{pmatrix} 0.80 & 0.20 \\ 0.10 & 0.90 \end{pmatrix} \begin{pmatrix} 0.05 \\ 0.10 \end{pmatrix} = 0.0106$$

To conclude, we estimate the belief $b_2(x_2) \triangleq p(x_2|x_3 = 1, x_6 = 1)$:

$$b_2(x_2) = \frac{1}{Z} m_{2\leftarrow d}(x_2) \circ m_{2\leftarrow c}(x_2) = \frac{1}{Z} \begin{pmatrix} 0.0091 \\ 0.0106 \end{pmatrix} \circ \begin{pmatrix} 0.99 \\ 0.01 \end{pmatrix} = \frac{1}{0.0091} \begin{pmatrix} 0.0090 \\ 0.0001 \end{pmatrix} =$$
$$= \begin{pmatrix} 0.99 \\ 0.01 \end{pmatrix}$$

Thus, the political leader won't probably won a lottery.

Problem 5 — *3 Points*

When posing the problem of binary image denoising as inference on a graphical model, we arrived at the following formulation:

$$x^\star = \arg\max_x \underbrace{\prod_i \exp^{-\alpha x_i} \prod_{j \text{ neighbor of } i} \exp^{-\beta x_i x_j}}_{a} \underbrace{\prod_i \exp^{-\gamma x_i y_i}}_{b}$$

where x was the clean image, y the observed noisy image, x_i, y_i the value of pixels at i, j and $x_i, y_i \in [+1, -1]$.

(a) where does this expression come from (develop probabilistic origin) ? Also, what's the name of parts (a) and (b) ?

(b) what's the meaning of α, β, γ ? what does it mean they are small or large ?

(c) which where the two assumptions made to arrive there ?

The term (a) is $p(x)$, the prior, that is a probability distribution over the set of all possible binary images which tells us the probability of a given image x. We know that normally binary images are not random noise and therefore there is a (possibly high) degree of smoothness in them, that is, one pixel tends to have the same value as its neighbors. In this prior model we are assuming (factorization) that the probability of a certain pixel given all other image pixels is the same as the probability given only its neighbors (Markovianity, Gibbs distribution). The degree of smoothness is controlled by parameter β. α expressess the preference for white or dark images.

$p(y|x)$ is the likelihood: the probability of a certain observed binary image y given the clean (but unknown) image x. The factorization implies that these probabilities are independent from one pixel w.r.t any other one (the noise is uncorrelated). The assumed degree of noise is modeled by γ.

We want to estimate $\arg\max_x p(x|y)$: the (jointly over all pixels) most probable clean image given the noisy one and a certain prior. The expression above comes from applying Bayes and removing the evidence term $p(y)$ which does not depend on x.

Problem 6 — *1 Point*

Can the former problem be solved with the min-cut/max-flow algorithm ? why or why not ?

Yes because 1) it's a problem of optimization of a pseudoquadratic boolean expression where 2) we can see the coefficients are all positive, thus suitable for min-cut. (optionally, add the prove)

Problem 7 — *1 Point*

What's false with respect to the graph-cuts algorithm ?

(a) does MAP inference on a graphical model

(b) repeatedly runs the max-flow/min-cut algorithm

(c) performs local "moves", eventually changing the label of many variables at each one

(d) can deal with any kind of potential functions

(e) it's iterative

(f) is false, only submodular potentials.