Master in Computer Vision Barcelona

UAB UOC UPC upf

Module: M2. Optimization and inference techniques for Computer Vision **Final exam**
Date: December 5th, 2019
Teachers: Juan F. Garamendi, Coloma Ballester, Karim Lekadir, Oriol Ramos, Joan Serrat **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

Juan F. Garamendi, 2 Points

Consider the following minimization problem:
Given the gray scale image $f \in L^{\infty}(\Omega)$ such that $u : \Omega \to \mathbb{R}$, solve

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

Where

- $\Omega \in \mathbb{R}^2$ is a bounded open domain of the 2 dimensional euclidean space $\mathbb{R}^2$.
- $W^{1,2}(\Omega) = \{u \in L^2(\Omega);\ \nabla u \in L^2(\Omega)^2\}$
- $\lambda > 0$ is a given (i.e. known) parameter. $\lambda \in \mathbb{R}$
- $\mathbf{x} \in \Omega$ such that $\mathbf{x} = (x_1, x_2)$ is the spatial variable and ∇ is the gradient operator such that $\nabla u(\mathbf{x}) = (u_{x_1}, .., u_{x_d})$

(a) (0.25 points) Say in a few words which is the image processing solved.
The problem solved is the image restoration problem. The unknown image u represents a denoised version of the given image f

(b) (0.25 points) Describe in a few (but concise) words the role of parameter λ. How λ affects to the solution if we decrease its value?
The parameter lambda is a weighting parameter that trade-off the regularity term and the data fidelity term. Decreasing λ gives less weight to the data fidelity term, or what is the same, gives more weight to the regularity term, so the solution will be smoother than using a higher λ.

(c) Knowing that

$$\frac{dJ}{du} = \lambda(u - f) - \Delta u$$

where Δu denotes the laplacian of u

- (0.5 points) Write a gradient descent scheme at the function level. Use f as initial image to the scheme.

$$\left\{ \begin{array}{lcl} u^{[0]} & = & f \\ u^{[k+1]} & = & u^{[k]} - \tau(\lambda(u - f) - \Delta u) \end{array} \right.$$

  Stops when $|u^{[k+1]} - u^{[k]}| < \epsilon$. Choose τ small. upper scripts in brackets means iteration number,

- (0.5 points) Discretize the previous gradient descent and write a explicit gradient descent at the pixel level using the following discretization for the second derivative

$$\frac{\partial^2 u}{\partial x^2} \approx u_{i+1,j} - 2u_{i,j} + u_{i-1,j}$$

$$\frac{\partial^2 u}{\partial y^2} \approx u_{ij+1} - 2u_{ij} + u_{ij-1}$$

  Use homogeneous Von-Neumann boundary conditions.
  knowing that $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$

$$\left\{ \begin{array}{lclll} u_{i,j}^{[0]} & = & f & & \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i+1,j}^{[k]} + u_{ij+1}^{[k]} + u_{ij-1}^{[k]} + (\lambda-3)u_{ij}^{[k]} - \lambda f) & \text{if} & i = 0 \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i+1,j}^{[k]} + u_{i-1,j}^{[k]} + u_{ij+1}^{[k]} + (\lambda-3)u_{ij}^{[k]} - \lambda f) & \text{if} & j = 0 \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i+1,j}^{[k]} + u_{i-1,j}^{[k]} + u_{ij-1}^{[k]} + (\lambda-3)u_{ij}^{[k]} - \lambda f) & \text{if} & j = \text{last column} \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i-1,j}^{[k]} + u_{ij+1}^{[k]} + u_{ij-1}^{[k]} + (\lambda-3)u_{ij}^{[k]} - \lambda f) & \text{if} & i = \text{last row} \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i+1,j}^{[k]} + u_{ij+1}^{[k]} + (\lambda-2)u_{ij}^{[k]} - \lambda f) & \text{if} & i,j = \text{upper left corner} \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i+1,j}^{[k]} + u_{ij-1}^{[k]} + (\lambda-2)u_{ij}^{[k]} - \lambda f) & \text{if} & i,j = \text{upper right corner} \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i-1,j}^{[k]} + u_{ij-1}^{[k]} + (\lambda-2)u_{ij}^{[k]} - \lambda f) & \text{if} & i,j = \text{bottom left corner} \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i-1,j}^{[k]} + u_{ij-1}^{[k]} + (\lambda-2)u_{ij}^{[k]} - \lambda f) & \text{if} & i,j = \text{bottom right corner} \\ u_{i,j}^{[k+1]} & = & u_{i,j}^{[k]} - \tau(u_{i+1,j}^{[k]} + u_{i-1,j}^{[k]} + u_{ij+1}^{[k]} + u_{ij-1}^{[k]} + (\lambda-4)u_{ij}^{[k]} - \lambda f) & & \text{in another case} \end{array} \right.$$

  Stops when $|u^{[k+1]} - u^{[k]}| < \epsilon$. Choose τ small. upper scripts in brackets means iteration number,

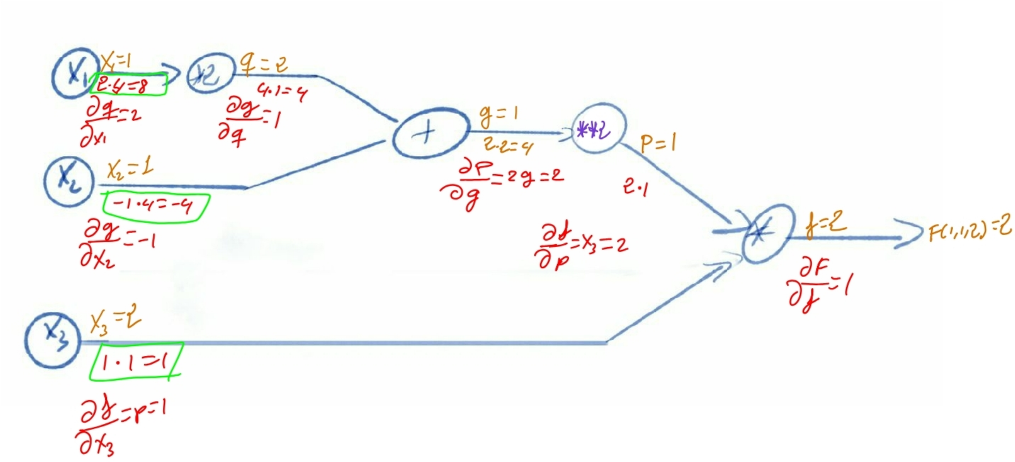
- (0.5 points) What happens if instead of taking f as initial image to the gradient descent scheme we take a black image? How this affects to the resolution of the problem? How will be the final image u compared with the previous scheme? (remember, we model the black color as 0, so the initial image is an image with all pixels values equal to zero).

  The problem is a convex problem, so the initial starting image only affects to the number of needed steps to get the minimum. If a black image is closer (or further) to the minimum (this will be depend on the image f) we will need less (or more) iterations, but we will get exactly the same denoised image.

Problem 2 *Juan F. Garamendi, 1 Point*

(a) (0.7 points) Compute the gradient at point $\bar{x} = (1, 1, 2)^T$ of the function $F(\bar{x}) = (2x_1 - x_2)^2 x_3$ using back-propagation. Write the flow graph and intermediate values for the forward passing as well as the back-propagation passing.

$F(x_1,x_2,x_3) = (2x_1 - x_2)^2 \cdot x_3$ at $(1,1,2)^T$.



Small pieces:

$q = 2x_1$; $g = q - x_2$ $\quad p = g^2 \quad f = p \cdot x_3$

Partial derivatives of the small pieces

$\frac{\partial q}{\partial x_1} = 2 \quad \frac{\partial g}{\partial q} = 1 \quad \frac{\partial p}{\partial g} = 2g \quad \frac{\partial f}{\partial p} = x_3$

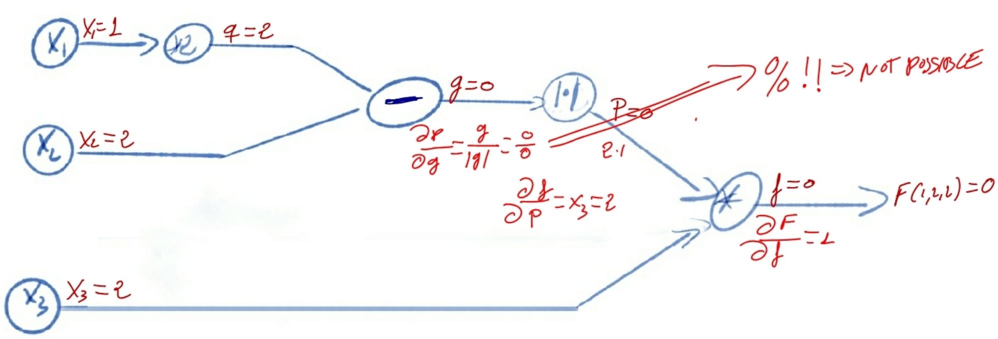
$\frac{\partial g}{\partial x_2} = -1 \quad \frac{\partial f}{\partial x_3} = p$

So $\nabla F = (8, -4, 1)^T$

(b) (0.15 points) What happens if we try to compute the gradient at point $\bar{x} = (1, 2, 2)^T$ of the function $F(\bar{x}) = |2x_1 - x_2|x_3$ using back-propagation? Will be any difference if we compute the gradient by hand or using some numerical method? (just if it is needed, remember that for $y = |x|$ the derivative is $y' = \frac{x}{|x|}$)

At point $\bar{x} = (1, 2, 2)^T$, $2x_1 - x_2 = 0$, so it is not derivable at that point and this fact is independent of which method you use for computing the derivative. Just as example, let see what happens if you try to do it by back-propagation.

$f(x) = |2x_1 - x_2| \cdot x_3$ $\quad \nabla f(x)$ at $(1,2,2)$



Small pieces:

$q = 2x_1$; $g = q - x_2$ $\quad P = |g|$ $\quad f = P \cdot x_3$

Partial derivatives of the small pieces

$\frac{\partial q}{\partial x_1} = 2$ $\quad \frac{\partial g}{\partial q} = 1$ $\quad \frac{\partial P}{\partial g} = \frac{g}{|g|}$ $\quad \frac{\partial f}{\partial P} = x_3$

$\frac{\partial g}{\partial x_2} = -1$ $\quad \frac{\partial f}{\partial x_3} = P$

(c) (0.15 points) Related with the question (b), What if for function from question (b) we try to compute the gradient at point from question (a)? Can we compute the derivative? (just say in a few words, you do not have to compute the gradient using back-propagation).

At point $\bar{x} = (1,1,2)^T$, $2x_1 - x_2 \neq 0$, so we do not have problems computing the derivative because of the absolute value.

Problem 3

Coloma Ballester *0.75 Points*

Consider the constrained minimization problem

$$\begin{aligned} \min_{x_1,x_2} \quad & -x_1 + 3x_2 \\ \text{subject to} \quad & \tfrac{1}{2}x_1 - x_2 + 2 \geq 0 \\ & x_1 - 3 \leq 0, \\ & x_1 + x_2 + 3 \geq 0. \end{aligned}$$

(a) Sketch the set of constraints of the problem. Is it a convex set? (0.2 *points*)

This is a problem of the form $\min_{\mathbf{x} \in C} f(\mathbf{x})$, where C is the convex set given by the closed triangle of the following figure:

(b) Write the KKT optimality conditions for the problem. (0.4 *points*)

The Lagrange dual function associated to the problem is

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2) = -x_1 + 3x_2 - \lambda_1(\frac{1}{2}x_1 - x_2 + 2) - \lambda_2(3 - x_1) - \lambda_3(x_1 + x_2 + 3)$$

which is convex with respect to x_1, x_2 (for λ_1, λ_2 fixed), and linear (thus concave) with respect to λ_1, λ_2 (for x_1, x_2 fixed). Thus, the KKT optimality conditions hold and are necessary and sufficient contditions of saddle point. The KKT optimality conditions are

$$\begin{cases} \nabla_x \mathcal{L}(x,\lambda_1,\lambda_2) = \begin{bmatrix} -1 \\ 3 \end{bmatrix} + \lambda_1 \begin{bmatrix} -\frac{1}{2} \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \lambda_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \frac{1}{2}x_1 - x_2 + 2 \geq 0 \\ 3 - x_1 \geq 0 \\ x_1 + x_2 + 3 \geq 0 \\ \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \\ \lambda_1(\frac{1}{2}x_1 - x_2 + 2) = 0 \\ \lambda_2(3 - x_1) = 0 \\ \lambda_3(x_1 + x_2 + 3) = 0. \end{cases}$$

(c) Check if any of the points $(x_1, x_2) = (0, -3)$ and $(x_1, x_2) = (3, -6)$ could be the solution of the problem using the KKT conditions. (0.15 *points*)

Both points belong to the feasible set C but only the second one, $(3, -6)$, satisfies the KKT optimality conditions. It is the solution of the problem as, in this case, they also are sufficient conditions.

Problem 4 Coloma Ballester *1.25 Points*

Let us consider the vectors $\mathbf{b}, \mathbf{d} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^m$, the $m \times n$ real matrix A, the real constants $\lambda, \mu > 0$, and the following minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{b}\| + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle$$

for $\mathbf{x} \in \mathbb{R}^n$.

(a) Is this a convex function? Why? (0.2 *points*)

Yes, the objective function, $\|\mathbf{x} - \mathbf{b}\| + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle$, is convex as it consists of the sum of convex functions: the composition of a norm and a linear function (both convex functions), plus a quadratic function which is twice differentiable with Hessian $2\lambda A^t A$ which is a positive definite matrix, plus a linear function.

(b) Describe why we can not use a gradient descent method to solve it. (0.15 *points*)

Because the objective function is not differentiable. Indeed, its first term is not differentiable.

(c) Write an equivalent min-max problem and the resulting iterations of a primal-dual algorithm to solve it. (0.5 *points*)

Using that $\|\mathbf{x} - \mathbf{b}\| = \max_{\xi \in C} \langle \mathbf{x} - \mathbf{b}, \xi \rangle$, where $C = \{\xi \in \mathbb{R}^n : \|\xi\| \leq 1\}$, we have that:

$$\|\mathbf{x} - \mathbf{b}\| + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle = \max_{\xi \in C} \left(\langle \mathbf{x} - \mathbf{b}, \xi \rangle + + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle \right),$$

Then:

$$\min_x \left(\|\mathbf{x} - \mathbf{b}\| + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle \right) = \min_x \max_{\xi \in C} \left(\langle \mathbf{x} - \mathbf{b}, \xi \rangle + + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle \right).$$

The function

$$\begin{aligned} \mathcal{L}(x, \xi) = \langle \mathbf{x} - \mathbf{b}, \xi \rangle + + \lambda \|\mathbf{A}\ \mathbf{x} - \mathbf{c}\|^2 + \mu \langle \mathbf{x}, \mathbf{d} \rangle &= \langle \mathbf{x} - \mathbf{b}, \xi \rangle + \lambda \langle \mathbf{A}\ \mathbf{x} - \mathbf{c}, \mathbf{A}\ \mathbf{x} - \mathbf{c} \rangle + \mu \langle \mathbf{x}, \mathbf{d} \rangle \\ &= \langle \mathbf{x}, \xi \rangle - \langle \mathbf{b}, \xi \rangle + \lambda \langle \mathbf{A}\ \mathbf{x} - \mathbf{c}, \mathbf{A}\ \mathbf{x} - \mathbf{c} \rangle + \mu \langle \mathbf{x}, \mathbf{d} \rangle \end{aligned}$$

depending on the primal variables x and on the dual variables ξ, is convex with respect x (for each $\xi \in C$ fixed) (it is quadratic on x and its Hessian is equal to $2\lambda A^t A$, which is a positive definite matrix) and concave with respect to ξ (for each $x \in \mathbb{R}^n$ fixed). Therefore, there exists a saddle point and the duality gap is zero. Thus, the original Primal problem, the Primal-Dual problem, and the Dual problem are three equivalent problems:

$$\min_x f(x) = \min_x \max_{\xi \in C} \mathcal{L}(x, \xi) = \max_{\xi \in C} \min_x \mathcal{L}(x, \xi) = \max_{\xi \in C} g_D(\xi)$$

The Primal-Dual problem is solved by an iterative Primal-Dual algorithm where we start from an initial (x^0, ξ^0), and alternate a projected gradient ascent step for the variable ξ, and gradient descent step for the variable x:

$$\xi^{k+1} = P_C(\xi^k + \tau \nabla_\xi \mathcal{L}(x^k, \xi^k))$$

$$x^{k+1} = x^k - \theta \nabla_x \mathcal{L}(x^k, \xi^{k+1}).$$

Here, $P_C(v) = \frac{v}{\max\{1, \|v\|\}}$ is a projector over C (for any vector vector $v \in \mathbb{R}^n$).
In our case, the 'partial gradients' of $\mathcal{L}$, denoted by $\nabla_x \mathcal{L}$ and $\nabla_\xi \mathcal{L}$ with respect to x and ξ, respectively, are given by

$$\nabla_x \mathcal{L}(x, \xi) = \xi + 2\lambda A^t(Ax - c) + \mu d$$

$$\nabla_\xi \mathcal{L}(x, \xi) = x - b$$

Finally, we find a solution by iterating the following update equations

$$\xi^{k+1} = P_C(\xi^k + \tau(x^k - b))$$

$$x^{k+1} = x^k - \theta \left(\xi^k + 2\lambda A^t(Ax^k - c) + \mu d \right),$$

(d) How the dual function is defined? Outline the dual algorithm to solve this problem? (0.4 *points*)

The dual function is $g_D(\xi) = \mathcal{L}(x_0(\xi), \xi)$, where

$$x_0(\xi) = \arg\min_x \mathcal{L}(x, \xi) = \arg\min_x \left(\langle \mathbf{x}, \xi \rangle - \langle \mathbf{b}, \xi \rangle + \lambda \langle \mathbf{A}\ \mathbf{x} - \mathbf{c}, \mathbf{A}\ \mathbf{x} - \mathbf{c} \rangle + \mu \langle \mathbf{x}, \mathbf{d} \rangle \right).$$

As it is a convex function on the whole space $\mathbb{R}^n$, the minimizer $x_0(\xi)$ is the solution of the necessary and sufficient condition $\nabla_x \mathcal{L}(x, \xi) = 0$, which is

$$x_0(\xi) = (A^t A)^{-1} \left(A^t c + \frac{1}{2} \lambda (\xi + \mu d) \right)$$

(assuming that $A^t A$ is invertible). Substituting $x_0(\xi)$ one obtains the dual function:

$$g_D(\xi) = \mathcal{L}(x_0(\xi), \xi) = ()$$

Finally the dual problem is

$$\max_{\xi \in C} g_D(\xi) = \max_{\xi \in C} ()$$

(which is a quadratic problem with concave objective function and with convex constraints, where we have eliminated the primal variable, and therefore could be solved with a projected gradient ascent).

Problem 5 Karim Lekadir *0.5 Points*

(a) Why is the Chan-Vese segmentation method also called "Active Contours Without Edges"?

Solution: Because it ignores edges and instead searches for coherent regions within the image. The edges are implicitly represented using a level set function.

(b) What is an appropriate choice for the initial level set function? Justify.

Solution: Checkboard shapes (2D sinusoidal) have been shown to converge more rapidly than other level set types (e.g. cone).

Problem 6 J.Serrat *0.5 Points*

We saw that binary image denoising could be modeled as a problem of maximum a posteriori inference over the graphical model of figure 1. The goal then was

$$\arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})$$

Which of the following statements are *false* ? (0.5 points if answer is exact)

(a) $p(\mathbf{y}|\mathbf{x})$ is the prior, $p(\mathbf{x})$ the likelihood

(b) $\mathbf{x}$ is the clean image (the pixels) and $\mathbf{y}$ the noisy one

(c) the order of the model is 2, also called pairwise

(d) the solution for the former equation is called the "*max-marginals*"

(e) $p(\mathbf{y})$ is the evidence

(f) in order to get a tractable expression for $p(\mathbf{x})$, we want the noise model to be independent (uncorrelated) and identically distributed

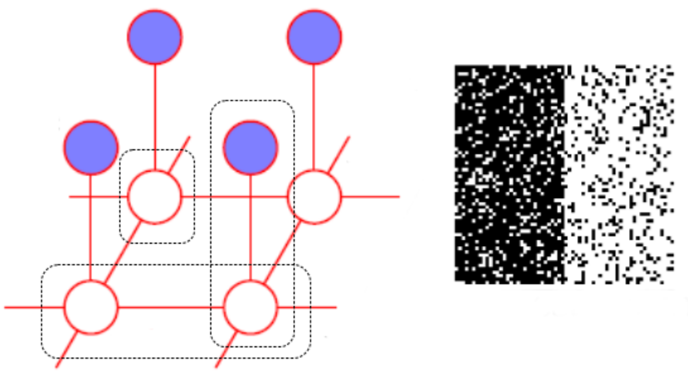


Figure 1: graphical model for binary image denoising

(a), (d), (f)

Problem 7 J.Serrat *0.5 Points*

The problem of dense disparity estimation from a pair of stero images I_L, I_R can be posed also as inference on a graphical model. We arrived at the following formulation:

$$\hat{D} = \arg\max_{D} p(D|I_L, I_R) = \arg\max_{D} p(I_L, I_R|D)\, p(D)$$

with for instance

$$\begin{aligned} p(I_L, I_R|D) &= \prod_i \exp[-(I_L(i) - I_R(i - D(i))] \\ p(D) &\propto \exp[-\sum_{i,j \in \mathrm{Ne}_i} |D(i) - D(j)|] \end{aligned}$$

Then, what is *false* again ? (one or more choices, 0.5 points if answer is exact)

(a) the likelihood expresses our preference for planar surfaces perpendicular to the viewing direction

(b) the prior says corresponding pixels have similar intensity

(c) $I_L(i) - I_R(i - D(i))$ will work worse than averaging differences over windows centered at i and $i - D(i)$ in I_L, I_R respectively

(d) it is convenient that prior and likelihood factorize and this actually happens

(e) $|D(i) - D(j)|$ may penalize too much the disparity differences at object borders

(f) one solution to the previous problem is to lower bound this value

(a), (b), (f)

Problem 8 J.Serrat *0.5 Points if (a) and (b) are correct*

Consider the graphical model of figure 2 where observations are binary images $16 \times 8 = 128$ pixels, that is, $x_i \in \{0, 1\}^{128}$, $y_i \in Y = \{\mathrm{a}, \mathrm{b} \ldots \mathrm{z}\}$ (26 lowercase letters), $x = (x_1 \ldots x_9), y = (y_1 \ldots y_9)$.

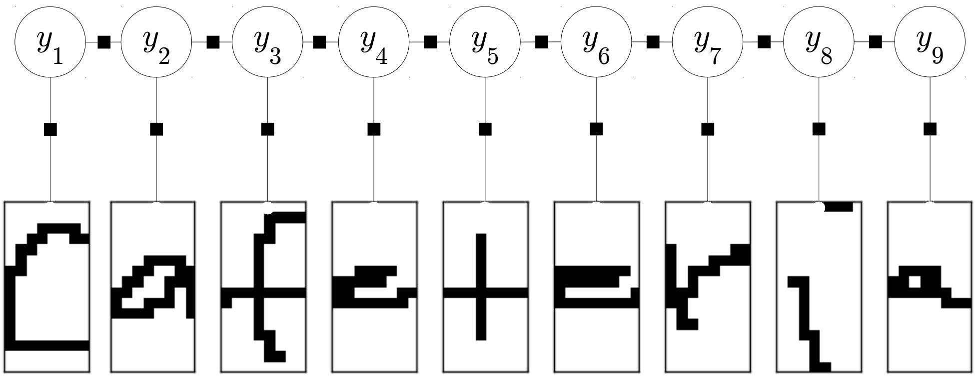


Figure 2

We want to learn w to later infer a word from a series of binary images of letters as

$$
\begin{aligned}
y^\star &= \underset{y \in \mathcal{Y}}{\arg\max} \ \langle w, \psi(x, y) \rangle \\
&= \underset{y \in Y^9}{\arg\max} \sum_{i=} \sum_{p=} \sum_{j=} \sum_{k=} w_{pjk}\, x_{ijk} \ + \sum_{l=} \sum_{p=} \sum_{q=} w_{pq}\, \mathbf{1}_{y_l = \mathrm{p},\, y_{l+1} = \mathrm{q}}
\end{aligned}
$$

where $\mathbf{1}_{y_i = \mathrm{p},\, y_{i+1} = \mathrm{q}}$ evaluates to 1 if $y_i = \mathrm{p}$ and $y_{i+1} = \mathrm{q}$. In this context,

(a) What are the initial and final values of each index in the summatories ?

(b) What would be if we apply the technique of two-stage training ?

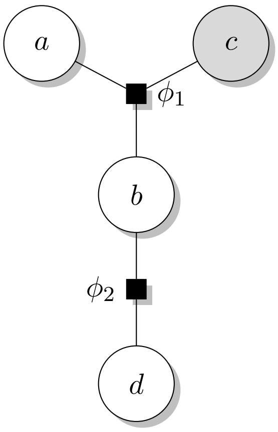
Answer right here :

	i	j	k	l	p	q
(a)						
(b)						

$$
\begin{aligned}
y^\star &= \operatorname*{arg\,max}_{y \in \mathcal{Y}} \langle\, w, \psi(x, y) \,\rangle \\
&= \operatorname*{arg\,max}_{y \in Y^9} \sum_{i=1}^{9} \sum_{p=\mathrm{a}}^{\mathrm{z}} \sum_{j=1}^{16} \sum_{k=1}^{8} w_{pjk}\, x_{ijk} \;+\; \sum_{i=1}^{8} \sum_{p=\mathrm{a}}^{\mathrm{z}} \sum_{q=\mathrm{a}}^{\mathrm{z}} w_{pq}\, \mathbf{1}_{y_i=\mathrm{p},\, y_{i+1}=\mathrm{q}}
\end{aligned}
$$

Problem 9 *Oriol Ramos Terrades, 1 Points*

Given the following factor graph:



$$\phi_1(a, b=0, c) = \begin{pmatrix} 10 & 5 \\ 20 & 3 \end{pmatrix}$$

$$\phi_1(a, b=1, c) = \begin{pmatrix} 4 & 7 \\ 2 & 5 \end{pmatrix}$$

$$\phi_2(b, d) = \begin{pmatrix} 3 & 8 \\ 4 & 9 \end{pmatrix}$$

with factors ϕ_1 and ϕ_2 defined above, compute:

a) The message sent by factor ϕ_2 to variable b: $m_{b\leftarrow 2}(b)$. Assume that d has not been observed [0.25 points].

Solution:

$$m_{b\leftarrow 2}(b) = \int \phi_2(b, d) m_{d\rightarrow 2}(d) dd = \begin{pmatrix} 3 & 4 \\ 8 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 17 \end{pmatrix}$$

b) The message sent by factor ϕ_1 to variable b: $m_{b\leftarrow 1}(b)$. Assume that c has been observed as $c = 1$ but a, not [0.5 points].

Solution:

$$m_{b\leftarrow 1}(0) = \int \phi_1(a, 0, c) m_{a\rightarrow 1}(a) m_{c\rightarrow 1}(c) da dc = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 10 & 5 \\ 20 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 23$$

$$m_{b\leftarrow 1}(1) = \int \phi_1(a, 1, c) m_{a\rightarrow 1}(a) m_{c\rightarrow 1}(c) da dc = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 7 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 7$$

c) The belief of b: $b(b)$ [0.25 points].

Solution:

$$b(b) = \frac{1}{Z_b} m_{b\leftarrow 1} \circ m_{b\leftarrow 2} = \frac{1}{Z_b} \begin{pmatrix} 7 \\ 17 \end{pmatrix} \circ \begin{pmatrix} 23 \\ 7 \end{pmatrix} = \frac{1}{280} \begin{pmatrix} 161 \\ 119 \end{pmatrix} = \begin{pmatrix} 0.575 \\ 0.425 \end{pmatrix}$$

where Z_b is the partition fucntion associated to belief b and is computed as: $Z_b = 161 + 119 = 280$.

Problem 10 *Oriol Ramos Terrades, 0.5 Points*

Say whether the next statements are true (**T**) or false (**F**) [Correct: +1/8, Incorrect: -1/8, unanswered: 0 points].

a) Belief Propagation (BP) infers exact marginals in undirected graphs.

b) An initial distribution is said to be *stationary* if, in the Markov chain specified by this initial distribution, the conditional distribution of x_{n+1} given x_n depend on n.

c) The Importance sampling method generates unbiased estimators.

d) Be $u \sim U(0,1)$, a random value uniformly drawn from the interval $[0,1]$, in the Metropolis-Hasting algorithm, we accept a new sample y if $u \leq \min\left\{1, \frac{h(y)q(y,x)}{h(x)q(x,y)}\right\}$.

Solution:

a) False.

b) False.

c) True.

d) True.