Master in Computer Vision
Barcelona
UAB UOC UPC upf.

Module: M2. Optimization and inference techniques for Computer Vision
Teachers: Juan F. Garamendi, Pablo Arias, Coloma Ballester, Karim Lekadir, Oriol Ramos
Final exam. Date: December 2nd, 2021 **Time: 2h30min**

- All sheets of paper should have your full name.
- The answer to all questions should be accurate and concise.
- You are allowed to use your notes, slides, papers, books, etc. The corresponding pdfs can be displayed on the PC screen but you are not allowed to use the internet nor any keyboard.
- You can answer the exercises one after another (in the order you choose) without leaving unnecessary blank spaces. Please use a horizontal continuous line to delimit each exercise and before the end of the exam cross out the extra text or calculations that do not make part of the final answer.
- If you finish before the 2 hours allocated for the exam, you will have to wait until the end when you will be informed that the task is opened for submissions.
- The exam delivery should be in .pdf format, uploaded to the task.
- Change the name of the file that Adobe Scan generates; the new name will be: Name_Surname.pdf

---

Problem 1

Juan F. Garamendi, 1.5 Points

Say whether the next statements are true (**T**) or false (**F**) [Correct: +0.25, Incorrect: -0.25, unanswered: 0 points].

(a) Given the image $f \in L^\infty(\Omega)$ finding u and v such that

$$u^* = \arg\min \left\{ \int_\Omega |\nabla u|^2 dxdy + \frac{1}{2\lambda} \int_\Omega |u - f|^2 dxdy \right\}$$

$$v^* = \arg\min \left\{ \int_\Omega |\nabla v| dxdy + \frac{1}{2\lambda} \int_\Omega |v - f|^2 dxdy \right\}$$

$u^* = v^*$ is satisfied

(b) Let $J(u)$ be a convex energy functional that we are minimizing on a convex set. If we use a gradient descent scheme to minimize it, the found solution u satisfies that $\nabla u = 0$

(c) For a given cost function $F(\bar{u}) = F(x_1, ..., x_N)$, If we use a gradient descent scheme to minimize it, the found solution $\bar{u}$ satisfies $\nabla F(\bar{u}) \simeq 0$.

(d) For a given cost function $F(\bar{u}) = F(x_1, ..., x_N)$, If we use the following iterative scheme (until convergence)

$$\begin{cases} \bar{u}^{[0]} & = & \bar{u}_0 \\ \bar{u}^{[k+1]} & = & \bar{u}^{[k]} - \tau \nabla F \end{cases}$$

the found solution $\bar{u}^{[k+1]}$ satisfies $\bar{u}^{[k+1]} \simeq 0$

(e) For a given cost function $F(\bar{u}) = F(x_1, ..., x_N)$, If we use the following iterative scheme (until convergence)

$$\begin{cases} \bar{u}^{[0]} & = & \bar{u}_0 \\ \bar{u}^{[k+1]} & = & \bar{u}^{[k]} - \tau \nabla F \end{cases}$$

the found solution $\bar{u}^{[k+1]}$ satisfies $\bar{u}^{[k+1]} \simeq \bar{u}^{[k]}$

(f) Given the image $f \in L^{\infty}(\Omega)$ finding u such that

$$u = \arg\min \left\{ \int_{\Omega} |\nabla u|^2 dxdy + \frac{1}{2\lambda} \int_{\Omega} |u - f|^2 dxdy \right\}$$

$u = f$ is satisfied

(a) False

(b) False

(c) True

(d) False

(e) True

(f) False

Problem 2 *Juan F. Garamendi, 1.5 Point*

For $\bar{v} \in \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$, Compute the gradient of function $F(\bar{v}) = |\bar{v}|^2$ at point $\bar{x}$ using back-propagation. Write the flow graph and intermediate values for the forward passing as well as the back-propagation passing. $|\cdot|$ stands for standard euclidean norm.

Solution $\nabla |\bar{v}|^2$ at $\bar{x}$ is $2\bar{x}$. Graph can be seen in figure 1

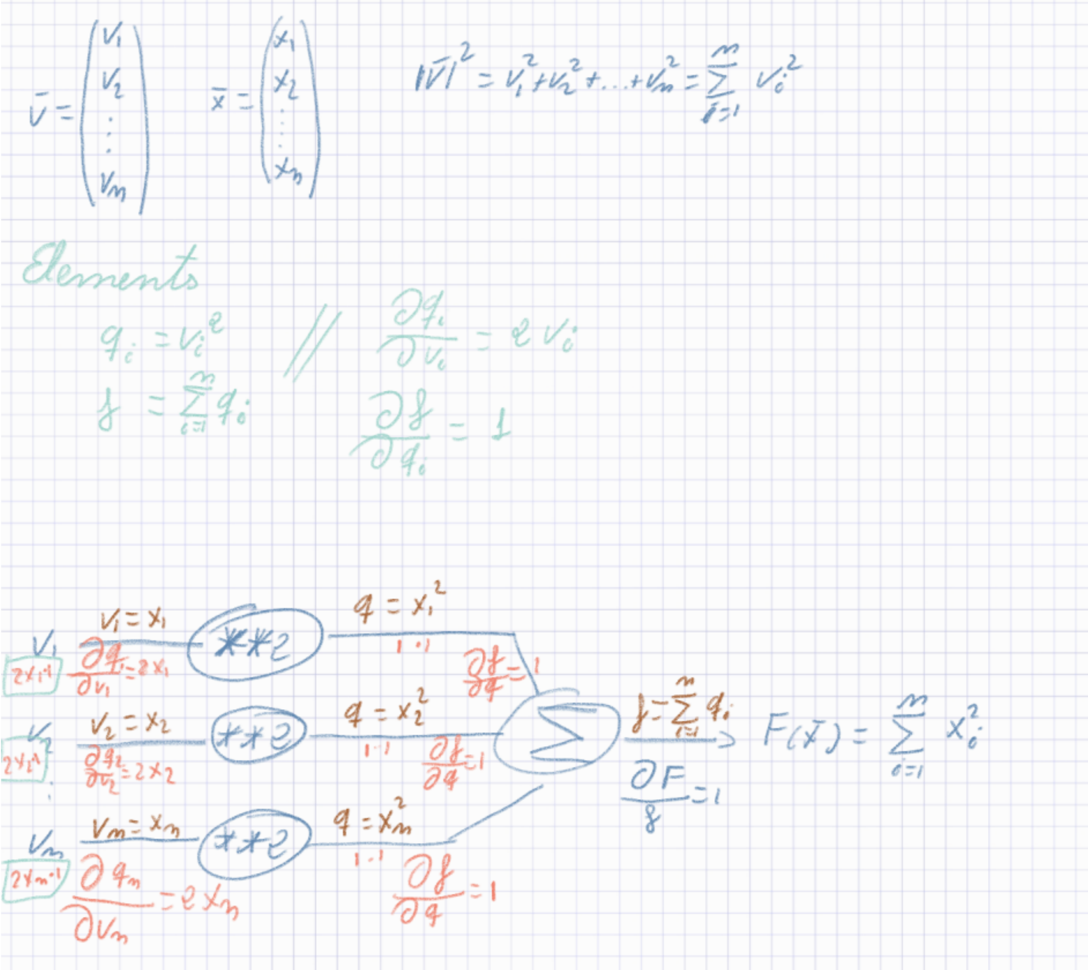


Figure 1: Solution for problem 3

Problem 3 Pablo Arias *0.7 Points*

Consider the following constrained optimization problem for $f : \mathbb{R}^2 \to \mathbb{R}$.

$$(\mathcal{P}) \begin{cases} \text{minimize} & f(x_1, x_2), \\ \text{subject to} & x_1 \geq 0, \\ & x_2 \geq 0. \end{cases}$$

All we know about f is that it is a *convex function*, and that $\mathrm{dom}(f) = \mathbb{R}^2$. We would like to know if solutions to $\mathcal{P}$ exist. By a solution to $\mathcal{P}$ we mean a minimizer of f in the feasible set. Answer the following questions providing a justification for your answers.

(a) Is the optimization problem $\mathcal{P}$ a convex problem? [0.1p]

(b) Could it be that there is not solution to problem $\mathcal{P}$? If so, provide an example. [0.2p]

(c) Could it be that there are multiple solutions to problem $\mathcal{P}$? If so, provide an example. [0.2p]

(d) Suppose that we add and additional constraint:

$$(\mathcal{P}') \begin{cases} \text{minimize} & f(x_1, x_2), \\ \text{subject to} & x_1 \geq 0, \\ & x_2 \geq 0, \\ & x_1^2 + x_2^2 \leq 10^6. \end{cases}$$

Could it be that the problem $\mathcal{P}'$ has no solution? If so, provide an example. [0.2p]

(a) An optimization problem is convex when f is a convex function and the feasible set is a convex set. We know that f is convex. The feasible set is given by the interesection of two halfspaces. Halfspaces are convex sets, and the intersection of convex sets is a convex set.

(b) Yes, there are functions f for which the problem has no solution. The feasible set a closed set, but it is not bounded, so we can not apply the Extreme Value Theorem. As an example, consider the function $f(x_1, x_2) = -x_1$. We have that $f(x_1, x_2) \to -\infty$ if $x_1 \to \infty$. Thus, there is no minimizer x^* such that $f(x) \geq f(x^*)$ for all feasible x.

(c) Yes, there are functions f for which the problem could have multiple solutions. Consider as an example the function $f(x_1, x_2) = 0$. In this case all feasible points are global minima. Another less trivial example is $f(x_1, x_2) = x_1$. In this case all points $(0, x_2)$ with $x_2 \leq 0$ are global minima.

(d) No, all convex functions f will have at least one minimizer of the feasible set. With the added constraint the feasible set is closed and bounded (it can be enclosed in a disk of radius 1000 centered at the origin). Since f is convex, it also continuous. Therefore we can now apply the Extreme Value Theorem: every continuous function on a closed and bounded set has a minimizer and a maximizer.

Problem 4

Pablo Arias *1.3 Points*

Consider the following constraint optimization problem, where d_1 and d_2 are parameters of a linear function f

$$(\mathcal{P}) \begin{cases} \text{minimize} & f(x_1, x_2) = d_1 x_1 + d_2 x_2, \\ \text{subject to} & x_2 \geq 4/x_1, \\ & x_2 \leq -x_1 + 5. \end{cases}$$

We are going to assume additionally that x_1 is positive, but to simplify the derivations we will not consider it explicitly in the set of constraints. Let $x^* = (x_1^*, x_2^*)$ a solution to $\mathcal{P}$.

(a) Sketch the feasible set. Is it a convex set? [0.2p]

(b) Write the KKT conditions for $\mathcal{P}$. [0.3p]

(c) Suppose that both constraints are *inactive* at x^*. Using the KKT conditions, show that this is only possible if $d_1 = d_2 = 0$. Which are the solutions in this case? [0.2p]

(d) Which are the possible solutions x^* if both constraints are *active* at x^*. [0.2p]

(e) Suppose that only the first constraint is *inactive* at x^*. Using the KKT conditions, show that this is only possible if $d_1 = d_2 = c$ for some negative number c. [0.2p]

(f) Suppose that only the second constraint is *inactive* at x^*. Show that [0.2p]

$$x_1^* = 2\sqrt{\frac{d_2}{d_1}}, \quad x_2^* = 2\sqrt{\frac{d_1}{d_2}}.$$

(g) Show that the previous case is only possible if $d_1/4 < d_2 < 4d_1$. [0.0p]

(a) A sketch of the feasible set is given in Figure 2. The first constraint $x_2 \geq 4/x_1$ can be described as the epigraph of the function $g(x_1) = 4x_1^{-1}$. Since the function g is convex its epigraph is also convex. The second constraint is a halfspace, and therefore is convex. Since both constraint sets are convex, their intersection is convex.

(b) We first write the problem following the sign convention for the constraints:

$$(\mathcal{P}) \begin{cases} \text{minimize} & f(x_1, x_2) = d_1 x_1 + d_2 x_2, \\ \text{subject to} & x_2 - 4/x_1 \geq 0, \\ & -x_1 - x_2 + 5 \geq 0, \end{cases}$$

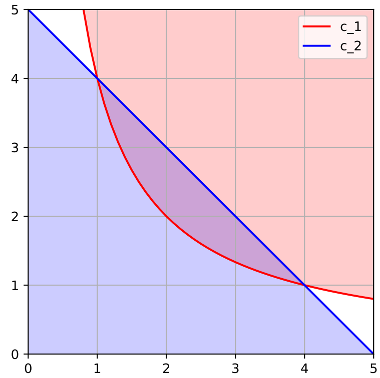


Figure 2: Feasible set. Shaded in red: the points that satisfy the first constraint, and in blue the points satisfying the second constraint. Their intersection is the feasible set.

The Lagrangian associated to our problem is given by

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2, \lambda_2) = d_1 x_1 + d_2 x_2 - \lambda_1(x_2 - 4/x_1) - \lambda_2(-x_1 - x_2 + 5).$$

The KKT conditions at a point x are:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x_1}(x, \lambda) = d_1 - \lambda_1 \dfrac{4}{x_1^2} + \lambda_2 = 0 & \text{stationarity 1} \\ \dfrac{\partial \mathcal{L}}{\partial x_2}(x, \lambda) = d_2 - \lambda_1 + \lambda_2 = 0 & \text{stationarity 2} \\ \\ x_2 - 4/x_1 \geq 0, & \text{primal feasibility 1} \\ -x_1 - x_2 + 5 \geq 0, & \text{primal feasibility 2} \\ \\ \lambda_1 \geq 0, & \text{dual feasibility 1} \\ \lambda_2 \geq 0, & \text{dual feasibility 2} \\ \\ \lambda_1(x_2 - 4/x_1) = 0, & \text{complementary slackness 1} \\ \lambda_2(-x_1 - x_2 + 5) = 0, & \text{complementary slackness 2} \end{cases}$$

(c) To simplify the notation, in the remaining parts we are going to write $x = (x_1, x_2)$ for the solution, instead of $x^* = (x_1^*, x_2^*)$.

If all constraints are inactive at x we have that $\lambda_1 = \lambda_2 = 0$. From the stationarity conditions if follows that $d_1 = d_2 = 0$. In this case, all feasible points are solutions.

(d) If the first two constraints are active at x then we have $\lambda_1, \lambda_2 > 0$. From the complementary slackness conditions we can conclude that

$$\begin{aligned} x_2 - 4/x_1 &= 0, \\ -x_1 - x_2 + 5 &= 0. \end{aligned}$$

In other words, the solution is going to be one of the intersection points of the red and blue curves in Figure 2. Which one will depend on the values of d_1 and d_2. To find these intersections, we substitute $x_2 = 4/x_1$ in the second equation:

$$-x_1 - 4/x_1 + 5 = 0 \quad \Leftrightarrow \quad x_1^2 - 5x_1 + 4 = 0.$$

Solving for this quadratic equation we have

$$x_1 = \frac{5 \pm \sqrt{5^2 - 4 \cdot 4}}{2} = \begin{cases} 1 \\ 4 \end{cases}.$$

We can now find the corresponding values of x_2 by using $x_2 = 4/x_1$. This gives $x = (1, 4)$ or $x = (4, 1)$.

(e) Since the first constraint is inactive, $\lambda_1 = 0$. Since the second constraint is active, we have that $\lambda_2 > 0$ and due to the complementary slackness condition, $x_2 = -x_1 + 5$. In other words, the solution is on the blue line in Figure 2.

Since $\lambda_1 = 0$, the stationarity conditions simplify to

$$\frac{\partial \mathcal{L}}{\partial x_1}(x, \lambda) = d_1 + \lambda_2 = 0, \quad \text{stationarity 1}$$
$$\frac{\partial \mathcal{L}}{\partial x_2}(x, \lambda) = d_2 + \lambda_2 = 0, \quad \text{stationarity 2}$$

From these equations we have that $d_1 = d_2 = -\lambda_2$. Therefore we set $c = -\lambda_2$. Since $\lambda_2 > 0$, $c < 0$.

(f) The second constraint is inactive and the first is active. Thus $\lambda_2 = 0$, and $\lambda_1 > 0$. Due to the complementary slackness condition we have that $x_2 = 4/x_1$. In other words, the solution is somewhere on the red curve in Figure 2.

The stationarity conditions simplify to

$$\frac{\partial \mathcal{L}}{\partial x_1}(x, \lambda) = d_1 - \lambda_1 \frac{4}{x_1^2} = 0 \quad \text{stationarity 1}$$
$$\frac{\partial \mathcal{L}}{\partial x_2}(x, \lambda) = d_2 - \lambda_1 = 0 \qquad \text{stationarity 2}$$

From the second equation we have that $d_2 = \lambda_1$, and substituting in the first we have that:

$$x_1^2 = 4\frac{d_2}{d_1} \quad \Rightarrow \quad x_1 = \pm 2\sqrt{\frac{d_2}{d_1}}.$$

Since $x_1 > 0$, the only solution is $x_1^2 = 2\sqrt{\frac{d_2}{d_1}}$. Using that $x_2 = 4/x_1$ we obtain $x_2 = 2\sqrt{\frac{d_1}{d_2}}$.

(g) In this exercise, we will find which are the conditions on d_1, d_2 so that the second constraint is inactive and the first constraint is active at x (i.e. solution on the red boundary of the feasible set). We know from the last part that $x = 2(\sqrt{d_2/d_1}, \sqrt{d_1/d_2})$. From Figure 2 we know that x_1 must be between 1 and 4. Therefore

$$1 < 2\sqrt{d_2/d_1} < 4 \quad \Rightarrow \quad \frac{d_1}{4} < d_2 < 4d_1.$$

Let us now show analytically that $1 \leq x_1 \leq 4$. From the primal feasibility condition of the 2nd constraint we know that

$$-x_1 - x_2 + 5 > 0.$$

This time the inequality is strict, because we are assuming that the this restriction is inactive. We have that $x_2 = 4/x_1$ because the first constaint is active. Substituting we have

$$-x_1 - 4/x_1 + 5 > 0 \quad \Rightarrow \quad x_1^2 - 5x_1 + 4 < 0.$$

On the left hand side we have a quadratic polynomial facing up. We know from the previous parts that the roots are 1 and 4. Therefore the polynomial is negative if $1 < x < 4$.

Problem 5

Coloma Ballester *1.5 Points*

I. For $\mathbf{x} \in \mathbb{R}^n$, consider the following minimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left(\ \|\mathbf{x}\| + \lambda \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \ \right)$$

where $\mathbf{A}$ is a real matrix of size $m \times n$, $\mathbf{b} \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}$, $\lambda > 0$.

(a) Write an equivalent min-max problem. Does it have a solution? Why? (0.25 $p.$)

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

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

Then,

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

The function

$$\mathcal{F}(x, \xi) = \langle \mathbf{x}, \xi \rangle + \lambda \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \langle \mathbf{x}, \xi \rangle + \lambda \langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle$$

depending on the primal variables $\mathbf{x}$ and on the dual variables ξ, is convex with respect to $\mathbf{x}$ (for each $\xi \in C$ fixed) (it is quadratic on $\mathbf{x}$ and its Hessian on $\mathbf{x}$ is equal to $2\lambda A^t A$, which is a positive definite matrix) and concave with respect to ξ (it is linear on ξ, for each $\mathbf{x} \in \mathbb{R}^n$ 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:

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

(b) Write the resulting iterations of a primal-dual algorithm to solve it. (0.5 $p.$)

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, for all $k \geq 0$:

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

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

Here, $P_C(v) = \frac{v}{\max\{1, \|v\|\}}$ is a projector over C (for any vector vector $v \in \mathbb{R}^m$).

In our case, the 'partial gradients' of $\mathcal{F}$, denoted by $\nabla_x \mathcal{F}$ and $\nabla_\xi \mathcal{F}$ with respect to x and ξ, respectively, are given by

$$\nabla_\xi \mathcal{F}(x, \xi) = x$$

$$\nabla_x \mathcal{F}(x, \xi) = \xi + 2\lambda A^t (Ax - b)$$

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

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

$$x^{k+1} = x^k - \theta \left(\xi^k + 2\lambda A^t (Ax^k - b) \right).$$

II. For $\mathbf{x} \in \mathbb{R}^n$, consider the following minimization problem

$$\begin{aligned} &\min \tfrac{1}{2} \langle \mathbf{x}, \mathbf{x} \rangle \\ &\text{subject to } \mathbf{A}\mathbf{x} - \mathbf{b} = 0, \\ &\text{and } \mathbf{x} \geq \mathbf{0} \end{aligned}$$

where $\mathbf{A}$ is a real matrix of size $m \times n$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{x} \geq \mathbf{0}$ is (as always) a notation for all coordinates $x_i \geq 0$ for all $i = 1, 2, \ldots, n$.

(a) Write the problem as a min-max problem. (0.15 *p.*)

$$\min_{Ax-b=0,\, x\geq 0} \frac{1}{2}\langle x, x\rangle = \min_{x} \max_{\lambda\in\mathbb{R}^n,\lambda\geq 0,\mu\in\mathbb{R}^m} \underbrace{\left(\frac{1}{2}\langle x, x\rangle - \langle \lambda, x\rangle - \langle \mu, Ax-b\rangle\right)}_{=\mathcal{L}(x,\lambda,\mu)}.$$

(b) Define the duality gap and prove that is equal to zero in this case. (0.25 *p.*)

The duality gap is the difference

$$DG = \min_{x\in R^n} \max_{\lambda\geq 0,\, \mu\in R^m} \mathcal{L}(x,\lambda,\mu) - \max_{\lambda\geq 0,\mu\in\mathbb{R}^m} \min_{x\in R^n} \mathcal{L}(x,\lambda,\mu).$$

which is always ≥ 0. In our case, there exists a saddle point of $\mathcal{L}(x,\lambda,\mu)$ and $DG = 0$ because $\mathcal{L}$ is convex with respect to x (for each $\lambda \geq 0$, $\mu \in R^m$ fixed) (it is quadratic on x and its Hessian on x is equal to the identity matrix I of size $n \times n$, which is a positive definite matrix) and concave with respect to (λ, μ) (it is linear on λ, μ, for each $x \in \mathbb{R}^n$ fixed).

(c) Compute the dual function of the problem. (0.25 *p.*)

From (b), the original Primal problem, the Primal-Dual problem, and the Dual problem are three equivalent problems:

$$\min_{Ax-b=0,\, x\geq 0} \frac{1}{2}\langle x, x\rangle = \min_{x} \max_{\lambda\geq 0,\mu} \mathcal{L}(x,\lambda,\mu) = \max_{\lambda\geq 0,\mu} \left[\min_{x} \mathcal{L}(x,\lambda,\mu)\right] = \max_{\lambda\geq 0,\mu} g_D(\xi).$$

The dual function is $g_D(\lambda,\mu) = \mathcal{L}(x^*(\lambda,\mu),\lambda,\mu)$, where

$$x^*(\lambda,\mu) = \arg\min_{x} \mathcal{L}(x,\lambda,\mu) = \arg\min_{x} \frac{1}{2}\langle x, x\rangle - \langle \lambda, x\rangle - \langle A^t\mu, x\rangle + \langle \mu, b\rangle.$$

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

$$x^*(\lambda,\mu) = \lambda + A^t\mu.$$

Substituting $x^*(\xi)$ one obtains the dual function:

$$g_D(\lambda,\mu) = -\frac{1}{2}\langle \lambda + A^t\mu, \lambda + A^t\mu\rangle + \langle \mu, b\rangle.$$

(d) Write down the dual problem. (0.1 *p.*)

$$\max_{\lambda\in\mathbb{R}^n,\lambda\geq 0,\mu\in\mathbb{R}^m} -\frac{1}{2}\langle \lambda + A^t\mu, \lambda + A^t\mu\rangle + \langle \mu, b\rangle.$$

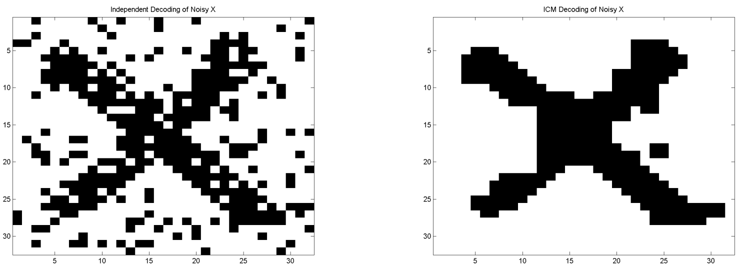
Problem 6 Karim Lekadir *0.5 Points*

(a) Given µ=0.1 in the Chan-Vese segmentation (for balancing the first term of the energy function), the obtained segmentation did not pick important details of the object's boundaries. What are the next values of µ would you test? Justify.

Solution: µ¡0.1 (e.g. µ=0.05 or µ= 0.01) to provide less weight to the first term, which is the contour smoothing term.

(b) In the Chan-Vese segmentation, the Heaviside function that is used to differentiate between the inside and outside of the regions is regularised. What is the purpose of this regularisation?

Solution: To make it differentiable and enable the minimisation of the objective function.



(a) Original Image (b) Denoised Image

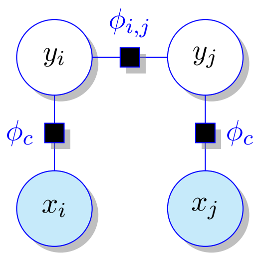
Figure 3: (a) Original binary image and (b) the denoised image

Problem 7 *Oriol Ramos Terrades, 1.5 Points*

Given the input binary image showed in Figure 3 (a), define a graphical model that will allow to obtained a denoised version of it like in Figure 3 (b). To do so, answer the following questions:

(a) Assume an image of size $N \times M$ pixel. Define the random variables of the proposed model and draw the associated graphical model. Do not forget to introduce the name of random variables, their domain and other needed notation to properly interpret the proposed graphical model [0.4 points].

Let $x_i \in \{0, 1\}$ the random variables corresponding to each image pixel and $y_i \in \{0, 1\}$ the hidden variables that will correspond to the denoised image. If we denote by i and j two adjacent pixels, according a 4 connectivity scheme, and $\mathcal{E}$ the set of adjacent pixels, the proposed graphical model is defined as:



(b) Write the joint distribution of the proposed method and determine which type of inference problem you must solve in order to obtain the denoised image [0.3 points].

The joint distribution is :

$$p(y|x) = \frac{1}{Z} \prod_n \phi_c(y_n, x_n) \prod_{i,j} \phi_{i,j}(y_i, y_j)$$

where Z is the partition function.

The denoised image is obtained after solving a Maximum a Posterior (MAP) problem of the previous joint distribution. In other words, the states of y that will maximize the MAP of the proposed joint model will correspond to the denoised image.

(c) Propose and justify candidate $\phi_{i,j}$ [0.3 points].

The simplest is the Potts model defined as:

$$\phi_{i,j} = \exp\big\{ -\theta 1_{\{y_i \neq y_j\}}(y_i, y_j) \big\}$$

for a positive real value θ that will control the denoising capacity of the model. High values of θ will have a "smoothing" effect.

(d) Can the partition function of the model, if exists, to be properly estimated for images of any size? If so, which inference algorithms can be applied to obtain it? If not, explain why [0.2 points].

Z can not properly estimated since for general undirected graphical models, as it is the case, it is a NP-hard problem. Moreover and although, is theoretical possible convert the proposed model to a junction tree, the complexity to estimate the beliefs for the bigger clusters make this estimation unfeasible in practice.

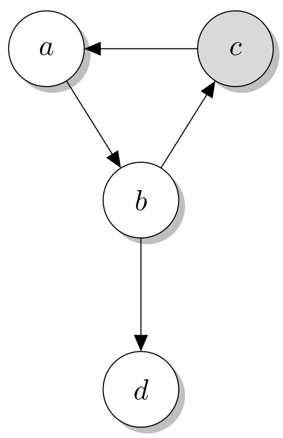
(e) Can a sampling method to be used to estimate a value of the partition function, if exists? If so, which of them, among those seen in the lectures, can be applied and why? If not, explain why [0.3 points].

We can apply normalized importance sampling and Gibbs sampling to estimate the partition function. For normalized importance sampling, the sampling distribution can simply be a Bernoulli distribution while for the Gibbs sampling we can rely on the conditional probability given by the Markov blanket.

We can not apply importance sampling because Z is unknown and Metropolis-Hasting is better suitable for continuous distributions.

Problem 8 *Oriol Ramos Terrades, 1.5 Points*

Given the following Bayesian network:



$p(a\|c)$	a	c
0.3	0	0
0.7	1	0
0.2	0	1
0.8	1	1

$p(c\|b)$	c	b
0.1	0	0
0.9	1	0
0.4	0	1
0.6	1	1

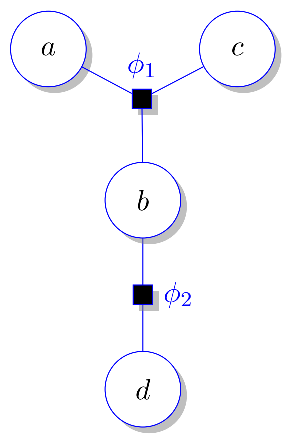
$p(d\|b)$	d	b
0.7	0	0
0.3	1	0
0.1	0	1
0.9	1	1

$p(b\|a)$	b	a
0.3	0	0
0.7	1	0
0.8	0	1
0.2	1	1

with prior and conditional probabilities as defined above:

a) Convert it, if possible, to a factor graph and define the joint probability [0.3 points].

Solution:



The joint probability is

$$p(a,b,c,d) = \frac{1}{Z} p(a|c)p(c|b)p(b|a)p(d|b),$$

being Z the partition function.

b) Define the corresponding factor functions, given the factor graph proposed in a) (no need to explicitly compute them) [0.2 points].

Solution: a possible definition of factor functions are:

$$\phi_1(a,b,c) = p(a|c)p(c|b)p(b|a)$$
$$\phi_2(b,d) = p(d|b)$$

c) Compute the message received by b from d. [0.4 points].

Solution:

The factor function $\phi_2(b,d)$ is :

$$\phi_2(b,d) = \begin{pmatrix} 0.7 & 0.1 \\ 0.3 & 0.9 \end{pmatrix}$$

since d is latent, the message sent by d to b is done through factor ϕ_2 as:

$$m_{b \leftarrow 2} = \int \phi_2(b,d)^t m_{d\to 2}(d) dd = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

d) Can belief propagation be used to infer exact marginals of a? If so, write the equations that will allow it assuming that c has been observed with value $c = 0$ (no need to explicitly compute messages). If not, explain why and propose an alternative way to estimate approximations of the marginals of a [0.6 points].

Solution:

Yes, since the proposed factor graph does not contain loops, we can choose a as root of a tree and send messages to it to estimate the exact marginal $p(a)$.

Given the current factorization, the message that will receive a is given by:

$$m_{a\leftarrow 1} = \int \phi_1(a,b,c) m_{b\to 1}(c) m_{c\to 1}(c) dbdc$$

Since $c = 0$ has been observed,

$$m_{c\to 1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$m_{b\to 1} = m_{b\leftarrow 2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The message received by a can be computed as:

$$m_{a\leftarrow 1}(0) = \begin{pmatrix} 1 & 0 \end{pmatrix} \phi_1(0,b,c)^t \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$m_{a\leftarrow 1}(1) = \begin{pmatrix} 1 & 0 \end{pmatrix} \phi_1(1,b,c)^t \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

where ϕ_1 has been defined by the conditional probabilities as in b).

Finally, The partition function of a is solely given message $m_{a\leftarrow 1}$ and therefore, the exact estimation of a is the normalized message.