Master in Computer Vision Barcelona

UAB UOC UPC upf

Module: M2. Optimization and inference techniques for Computer Vision **Final exam**
Date: November 30th, 2017
Teachers: Juan F. Garamendi, Coloma Ballester, 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, 1 Point

Let

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

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 euclidean space $\mathbb{R}^d$.
- $u \in \mathcal{V}$, $u : \Omega \rightarrow \mathbb{R}$ is a scalar function defined on Ω.
- $\mathbf{x} \in \Omega$ such that $\mathbf{x} = (x_1, .., x_d)$ 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 fundamental problem in calculus of variations.

(b) (0.25 points) Write the definition of the Gâteaux derivative of $J(u)$ (the directional derivative of J at u in the direction of a function $h(\mathbf{x})$).

(c) (0.5 points) Say in a few words what are the Euler-Lagrange Equations for a given convex functional.

Problem 2

Juan F. Garamendi, 1 Point

The OpenCV library (Open source Computer Vision Library) has a class called SVD. In the cv::SVD class reference documentation is said the following:

```
Class for computing Singular Value Decomposition of a floating-point matrix.
The singular Value Decomposition is used to blah, blah, blah, blah
```

(a) (0.5 points) Say at least two problems in which SVD is useful.

(b) (0.5 points) Say in a few words how to solve the problem (or at least compute an approximated solution to the problem) $A\bar{x} = b$ where A is a mxn matrix and $\bar{x}$ and $\bar{b}$ are vectors of size n and m, respectively, with $m > n$, and $\bar{x}$ is the unknown.

Problem 3 *Juan F. Garamendi, 1. Points*

Consider the following iterative scheme

- $\bar{x}^k \leftarrow S_i(\bar{x}^{k-1}, \bar{x}^k, \bar{b})$

used to solve the algebraic problem $\mathbf{A}\bar{x} = \bar{b}$, i.e., $\lim_{k\to\infty} \bar{x}^k = \bar{x}$, where $\mathbf{A}$ is a known matrix, $\bar{b}$ is a known vector, $\bar{x}$ is an unknown vector, S_i some function, super-index represents iteration number and $\bar{x}^0$ some initial value. Now consider two versions of S_i: S_1, S_2 with the following behaviour

(a) $e = \bar{x} - S_1(\bar{x}^1, \bar{x}^2, \bar{b})$, $\|e\|_\infty = 10^{-10}$, with e having only high frequencies.

(b) $e = \bar{x} - S_2(\bar{x}^1, \bar{x}^2, \bar{b})$, $\|e\|_\infty = 1000$, with e having only low frequencies.

(a) (0.5 points) Explain in few words which is the best iterative scheme (S_1 or S_2) for embedding it into a multigrid scheme.

(b) (0.5 points) Explain the relationship between Gauss-Seidel and Multigrid and in which moment of a multigrid alogrithm is used Gauss-Seidel. Is Gauss-Seidel the best scheme to use with multigrid?

Problem 4 Coloma Ballester *1. Points*

(a) (i) What is a convex set? (ii) Give an example of a convex set. (0.2 points)

(b) (i) What is a convex function (or a convex functional)? (ii) Give an example of a convex function $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ (that is, $f(\mathbf{x}) \in \mathbb{R}$ for any vector $\mathbf{x} \in \mathbb{R}^N$), and an example of a convex functional $J : V \longrightarrow \mathbb{R}$ (that is, $J(u) \in \mathbb{R}$ for any function $u \in V$, where V is a suitable space of functions). Give also an example of a non-convex functional (or function, if you prefer). (0.4 points)

(c) Consider the following constrained optimization problem (0.4 points)

$$(\mathcal{P}) \begin{cases} \min\limits_{x_1,x_2} f(x_1,x_2) = x_1 + x_2 \\ \text{subject to} \\ \quad 2 - x_1^2 - x_2^2 \geq 0 \\ \quad x_2 \geq 0 \end{cases}$$

(i) This is a problem of the form $\min\limits_{\mathbf{x}\in C} f(\mathbf{x})$. Draw a picture of the constraint set C.

(ii) What are the associated Karush-Kuhn-Tucker (KKT) optimality conditions?

Problem 5 Coloma Ballester *1 Points*

Choose and answer only one of the following two options:

Option 5.A. Let A be a $m \times n$ matrix, and $b \in \mathbb{R}^n$, $m, n \in \mathbb{N}$. Consider the problem of minimizing the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = \|Ax\|_{\mathbb{R}^m} + \frac{1}{2\lambda}\|x - b\|^2_{\mathbb{R}^n}$$

for $\lambda > 0$. The notation $\|\cdot\|_{\mathbb{R}^k}$ stands for the Euclidean norm in $\mathbb{R}^k$, for $k \in \mathbb{N}$. Note that f is not differentiable when $Ax = 0$.

(a) Write the problem

$$\min_x f(x) \qquad \text{(P)}$$

as a min-max problem. Define also the duality gap.
Hint: Remember the fact that $\|y\|_{\mathbb{R}^k} = \max_{\|\xi\|_{\mathbb{R}^k} \le 1} \langle y, \xi \rangle_{\mathbb{R}^k}$.

(b) What is the primal-dual problem for problem (P)? Are they equivalent problems?

(c) Define and compute the dual function and the dual problem of problem (P).

Option 5.B. Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Consider the problem

$$\begin{aligned} &\min \|x\|^2 \qquad \text{(P)} \\ &\text{subject to } Ax = b. \end{aligned}$$

(a) Write problem (P) as a min-max problem and define the duality gap.

(b) Define and compute the dual function of problem (P).

(c) Write down the dual problem.

Problem 6 *Joan Serrat, 0.5 Points*

The three main problems found when trying to learn the parameters of a graphical model in the probabilistic formulation are :
Problems:

(1) $Z(x^i, w)$ or $\mathbb{E}_{y \sim p(y|x^i,w)} \psi(x^i, y)$ impossible to calculate in practice

(2) N large and therefore we have to run belief propagation N times

(3) N small compared to number of parameters, causing overfitting

We saw that they could be overcomed by :

(a) regularization, assuming w follows a Gaussian distribution

(b) since $\psi(x, y)$ decomposes in factors, we can apply some inference method like belief propagation to compute it

(c) perform stochastic gradient descent

Now, which problem is solved by what ?

a) 1-a, 2-b, 3-c

b) 1-b, 2-c, 3-a

c) 1-c, 2-a, 3-b

d) 1-b, 2-a, 3-c

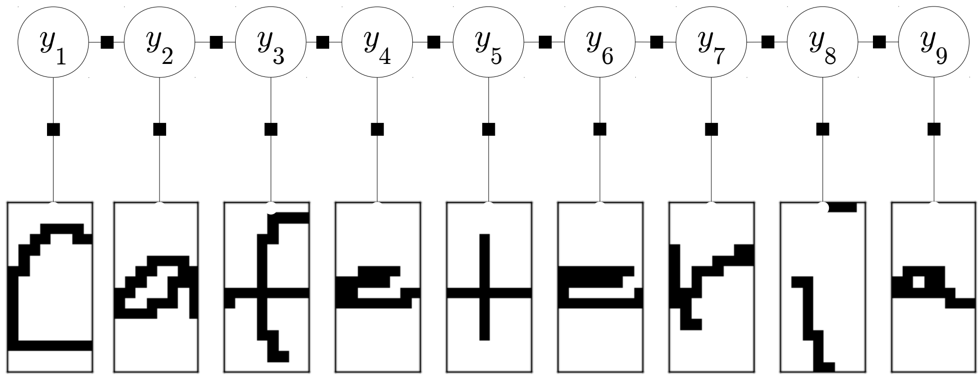


Figura 1

Problem 7 J.Serrat *0.5 Points*

Consider the graphical model of Figure 1 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), $x = (x_1 \ldots x_9), y = (y_1 \ldots y_9)$.

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

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

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, what's **false** ? Can be one or more.

a) the total number of unary parameters to learn is 26^2

b) there are $16{\times}8{\times}26$ pairwise parameters

c) $w_{p=\mathrm{a},q=\mathrm{b}}$ means the compatibility of labeling p as letter a and q as letter b, being p and q any pair of nodes in the chain

d) w_{pq} are the parameters of the prior term

Problem 8 *Joan Serrat, 0.5 Points*

The two-stage training is a technique (mark the true one)

a) to learn such that then makes inference faster

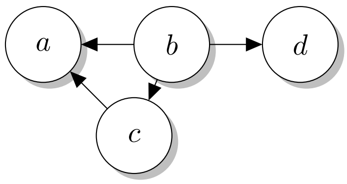
b) to learn a balanced set of unary and pairwise coefficients

c) that implies not learning at all the unary coefficients

d) to speed up learning only

Problem 9 *Oriol Ramos Terrades, 0.5 Points*

Given the following Bayesian network:



a) Write the joint distribution according to the conditional probabilities inferred from the Bayesian network.

b) Draw a factor graph derived from it.

Problem 10 *Oriol Ramos Terrades, 1 Point*

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

a) Belief propagation infer exact marginals in loopy graphical models.

b) The complexity of loopy belief propagation depends on the order of the highest clique.

c) Samples generated by the Metropolis-Hasting algorithm are always accepted.

d) Sampling methods provides exact inference on tree-based graphical models.