Master in Computer Vision Barcelona

UAB UOC UPC upf

Module: **M2. Optimization and inference techniques for Computer Vision**
nal exam
Date: ??, 2014
Teachers: ..., ... **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 *0.75 Points*

Consider the function $f : R^n \to R$ defined by $f(x) = \|Ax - b\|^2$, for $x \in R^n$, where A is a $m \times n$ matrix, and $b \in R^m$ $(m, n \in N)$. (The notation $\|\cdot\|$ stands for the Euclidean norm, and $\langle \cdot, \cdot \rangle$ stands for the scalar product.)

(a) Let $v \in R^n, v \neq 0$. Compute $D_v f(x)$, the directional derivative of f at the point x in the direction v.

(b) Use the result in (a) to compute $\nabla f(x)$.

(c) Which is the equation satisfied by a minimum of $f(x)$? (it is called the Euler-Lagrange equation).

(a) Given $v \in R^n, v \neq 0$, and $x \in R^n$:

$$D_v f(x) = \lim_{\epsilon \to 0+} \frac{f(x + \epsilon v) - f(x)}{\epsilon} = \lim_{\epsilon \to 0+} \frac{\|A(x + \epsilon v) - b\|_2^2 - \|Ax - b\|_2^2}{\epsilon} =$$

$$= \lim_{\epsilon \to 0+} \frac{\langle A(x + \epsilon v) - b, A(x + \epsilon v) - b \rangle - \langle Ax - b, Ax - b \rangle}{\epsilon}$$

$$= \lim_{\epsilon \to 0+} \frac{\langle Ax - b, Ax - b \rangle + \epsilon \langle Ax - b, Av \rangle + \epsilon \langle Av, Ax - b \rangle + \epsilon^2 \langle Av, Av \rangle - \langle Ax - b, Ax - b \rangle}{\epsilon}$$

$$= \lim_{\epsilon \to 0+} \frac{2\epsilon \langle Ax - b, Av \rangle + \epsilon^2 \langle Av, Av \rangle}{\epsilon} = 2\langle Ax - b, Av \rangle = \langle 2A^t(Ax - b), v \rangle$$

(b) We know

$$D_v f(x_0) = \langle \nabla f(x), v \rangle.$$

for all $v \in R^n$. We can extract from this the vector $\nabla f(x)$:

$$\nabla f(x) = 2A^t(Ax - b).$$

(c) The equations satisfied by the minimum are

$$\nabla f(x) = 0.$$

In our case, $2A^t(Ax - b) = 0$, that is, $A^tAx = A^tb$.

Problem 2 *1 Point*

Let A be a $m \times n$ matrix, and $b \in R^m$. Consider, for $x \in R^n$, the function $f(x) = \langle x, x \rangle$. Consider the problem (P) defined as

$$\begin{array}{l} \min f(x) \\ \text{subject to } Ax = b. \end{array}$$

(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.

(a) $Ax = b$ gives m equality constraints on x: $(Ax)_i = b_i$, $i = 1, \cdots, m$. Therefore, we introduce m Lagrange multipliers (or dual variables), $\nu_1, \dots, \nu_m$, and we construct the Lagrangian function $\mathcal{L}(x, \nu) = f(x) - \sum_{i=1}^{m} \nu_i((Ax)_i - b_i) = \langle x, x \rangle - \langle \nu, Ax - b \rangle$, where $\nu = (\nu_1, \dots, \nu_m)^t \in R^m$. Therefore

$$\min_{\text{subject to } Ax=b} \langle x, x \rangle = \min_{x \in R^n} \max_{\nu \in R^m} \mathcal{L}(x, \nu)$$

The duality gap is the difference

$$DG = \min_{x \in R^n} \max_{\nu \in R^m} \mathcal{L}(x, \nu) - \max_{\nu \in R^m} \min_{x \in R^n} \mathcal{L}(x, \nu).$$

(b) In our case, $DG = 0$ because $\mathcal{L}(x, \nu)$ is convex on x (f is convex and $d_i(x) = (Ax)_i - b_i$ are linear constraints) and $\mathcal{L}(x, \nu)$ is concave on ν. Therefore we can change min-max by max-min:

$$\min_{\text{subject to } Ax=b} \langle x, x \rangle = \min_{x \in R^n} \max_{\nu \in R^m} \mathcal{L}(x, \nu) = \max_{\nu \in R^m} \min_{x \in R^n} \mathcal{L}(x, \nu) = \max_{\nu \in R^m} g_D(\nu),$$

where $g_D(\nu) = \min_{x \in R^n} \mathcal{L}(x, \nu)$ is the dual function. We compute the dual function from $\nabla_x \mathcal{L}(x, \nu) = 0$ (indeed, $\mathcal{L}(x, \nu)$ is a quadratic function of x, and for each ν there is a unique minimizer $x_0(\nu)$. The minimizer is the solution of $\nabla_x \mathcal{L}(x, \nu) = 0$). In our case, $2x - A^t\nu = 0$, which gives $x_0(\nu) = \frac{1}{2}A^t\nu$. Then,

$$g_D(\nu) = \mathcal{L}(x_0(\nu), \nu) = -\frac{1}{4}\langle A^t\nu, A^t\nu \rangle + \langle \nu, b \rangle$$

(c) $\max_{\nu \in R^m} (-\frac{1}{4}\langle A^t\nu, A^t\nu \rangle + \langle \nu, b \rangle)$.

Problem 3 *0.75 Points*

Consider the following data fitting (or regression) problem: we are given a data set of N pairs $(t_1, y_1), \dots, (t_N, y_N)$, where $t_i \in R$, $y_i \in R$, $i = 1, \dots, N$, with $N > 10$. We would like to fit a function f to this known data set and assume that we know that there exist a functional relationship $y = f(t)$, with f modeled as the following parametric function

$$f(t) = x_1 + x_2t + x_3t^2$$

where $x_1, x_2, x_3 \in R$ are unknows.

(a) Explain the least squares solution to this problem of data fitting, used to determine the best parameters $\mathbf{x} = (x_1, x_2, x_3)$. Write down the expression of the energy $E(\mathbf{x})$ (or $E(x_1, x_2, x_3)$) to be minimized.

(b) Write down the normal equations associated to this problem.

(c) How could you determine the parameters $\mathbf{x} = (x_1, x_2, x_3)$ using the SVD or the pseudoinverse of the matrix associated to your problem? (Recall that SVD stands for Singular Value Decomposition of a matrix).

(a) When fitting the polynomial of degree 2, $f(t) = x_1 + x_2 t + x_3 t^2$, to the data $(t_1, y_1), \ldots, (t_N, y_N)$, the coeffients x_i are unknowns we have to determine. We talk about linear regression because although f is not a linear function of t, it is linear on the coeffients x_i.

There are more data than unknowns, $N > 3$. We compute the unknowns x_i by minimizing the sum of squared errors, between the predicted value $f(t)$ and the measured value, y, over the data:

$$\min_{x_0,x_1,x_2} \sum_{i=1}^{N} |f(t_i) - y_i|^2 = \min_{x_0,x_1,x_2} \sum_{i=1}^{N} \left|x_0 + x_1 t_i + x_2 t_i^2 - y_i\right|^2$$

Therefore, the energy to be minimized is $E(x_0, x_1, x_2) = \sum_{i=1}^{N} \left|x_0 + x_1 t_i + x_2 t_i^2 - y_i\right|^2$. We can write E and the above problem in matrix notation as

$$\min_x E(\mathbf{x}) = \min_x \|A\mathbf{x} - \mathbf{y}\|^2 \tag{1}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_N]^t \in R^N$, $\mathbf{x} = [x_0, x_1, x_2]^t \in R^3$ and

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_N & t_N^2 \end{bmatrix}$$

(b) The normal equations associated to the least squares problem (1) are $A^t A\mathbf{x} = A^t \mathbf{y}$.

(c) The solution of the normal equations with minimum norm is $\bar{\mathbf{x}} = A^+ \mathbf{y}$ where A^+ is the pseudoinverse of A. To compute the pseudoinverse, we use the SVD decomposition of A, which in turn is:

$$A = U \Sigma V^t$$

where U is a $N \times N$ orthogonal matrix, V is a 3×3 orthogonal matrix, and Σ is a rectangular $N \times 3$ diagonal matrix. The diagonal values of Σ, $\sigma_i > 0$, for $i = 1, \ldots, r$ are the sigular values. The number of singular values r is the rank of A (thus, $r \leqslant 3 = \min\{N, 3\}$).
From the SVD of A, we compute its pseudoinverse:

$$A^+ = V(\Sigma^t)^+ U^t$$

where the pseudo-inverse $(\Sigma^t)^+$ is a $3 \times N$ diagonal matrix with entries

$$\sigma_i^+ = \begin{cases} \frac{1}{\sigma_i} & \text{if } \sigma_i \neq 0 \\ 0 & \text{if } \sigma_i = 0, \end{cases}$$

for $i = 1, 2, 3$.