ENT 305A: Programming exercises

Laurent Pfeiffer, Inria-Saclay

September 2023

Exercice 6 (Logistic regression)**.** A *least-square problem* is an optimization problem of the form

$$\inf_{x \in \mathbb{R}^n} J(x), \quad \text{where: } J(x) = \sum_{i=1}^{m} g_i(x)^2,$$

and where $g : \mathbb{R}^n \to \mathbb{R}^m$ is a given mapping.

Figure 1 contains data of the population of the United States from 1790 to 2000. You will find the corresponding `data` file on the website of the class.

Year	Population
1790	3.93
1800	5.31
1810	7.24
1820	9.64
1830	12.86
1840	17.06
1850	23.19
1860	31.44
1870	38.56
1880	50.19
1890	62.98

Year	Population
1900	76.21
1910	92.23
1920	106.02
1930	123.20
1940	132.16
1950	151.33
1960	179.32
1970	203.30
1980	226.54
1990	248.71
2000	281.42

Figure 1: Population of the United States from 1790 to 2000, in millions

The following model, called *logistic model* is frequently used in population modelling:

$$P(t) = \frac{p}{1 + \exp(-\alpha(t - t_0))},$$

where $p > 0$, $\alpha > 0$ and $t_0 \in \mathbb{R}$ are three coefficients to be fixed.

1. Give an interpretation of the three coefficients. Draw a graph of P.
2. Formulate a least-square problem that allows to find out values of p, α and t_0 such that the corresponding time-function P models in a precise way the evolution of the american population.
3. Without appropriate initialization of the coefficients, AMPL may have difficulties to solve the problem (which is not convex). Propose a method that allows to suitably initialize p, α, and t_0.
4. Solve the problem with AMPL.

Solution: $p = 440.834$, $\alpha = 0.0216059$, $t_0 = 1976.63$.

Exercice 7 (Water distribution network)**.** We consider a water distribution network, made of a reservoir providing an agricultural area with water. The capacity of the reservoir (the maximal volume of water that can be stocked) is 1950 m^3, the minimal volume is 550 m^3. The initial volume is 1500 m^3.

The incoming flow (into the reservoir) varies over the year. It is denoted $Q(t)$, where $t = 1, \ldots, 12$ denotes the month of the year. It is expressed in m^3 in the following table:

t	1	2	3	4	5	6	7	8	9	10	11	12
$Q(t)$	386	346	416	713	1532	2000	1982	1431	780	476	450	420

The available agricultural area amounts to $A_{\max} = 200$ ha and can be divided into four cereals: cotton, wheat, rice, alfalfa. The revenue generated by each cereal is given in the following table, in euro per ha and per year:

	Cotton	Wheat	Rice	Alfalfa
Revenue	3500	700	2205	500

The needs for water of each cereal are given below (in m^3/ha), for each month of the year:

t	1	2	3	4	5	6	7	8	9	10	11	12
Cotton	3	3	3	2	4	5	5	4	5	0	0	0
Wheat	0	0	0	1	3	3	0	0	0.2	0	0	0
Rice	0	0	0	4.5	7	8	8	7	0.5	0	0	0
Alfalfa	0	0	0	0.5	2	3.5	3.5	2.5	0.5	0	0	0

Let $R(t)$ denote the amount of water withdrawn from the reservoir during the month t. Let $S(t)$ denote the amount of water at the end on the month t. Only a part $D(t) \leq R(t)$ of the withdrawn water is allocated to the agricultural area. Half of the allocated water gets lost and so the volume of water dedicated to the cultures amounts to $0.5D(t)$. Finally, the volume of water in the reservoir at the end of the year must be greater or equal to $S_f = 1400$ m^3.

We look for the areas x_1, x_2, x_3, and x_4 of each cereal (in ha) which maximize the annual revenue generated by their exploitation.

1. List the different optimization variables and parameters of the problem. Write the full optimization problem.
2. Solve the problem with AMPL. Some of the parameters have already been written in a data file on the website of the class.

Solution. Cotton: 114.87, wheat: 32.1295, rice: 20.8263, alfalfa: 32.1747.

Exercice 8 (Electrical grid)**.** We consider an electrical grid, composed of generation and consumption stations and transmission lines. We represent the grid with a graph made of N nodes (for the stations) and M edges (for the lines). An edge $e \in \{1, ..., M\}$ is characterized by an initial node $d(e) \in \{1, ..., N\}$, a terminal node $s(e) \in \{1, ..., N\}$, and a resistance $R_e > 0$. At each node $k \in \{1, ..., N\}$, a current of intensity J_k is injected (when a current is withdrawn from the grid, we have $J_k < 0$). The sum of all injected currents must necessarily be null, thus the parameters $(J_k)_{k=1,...,N}$ must satisfy

$$\sum_{k=1}^{N} J_k = 0.$$

We give below an example with 4 nodes and 5 edges:

Edge e	Initial node $d(e)$	Terminal node $s(e)$	Resistance R_e
1	1	2	1
2	1	4	1
3	2	4	1
4	1	3	1
5	3	4	1

We will take $J_1 = 3$, $J_2 = -1$, $J_3 = -1$, $J_4 = -1$.

The goal of the exercise is to predict how the electric current spreads over the grid. We denote by I_e the intensity of the current over the edge e, from $d(e)$ to $s(e)$. For any node k, we denote

$$\sigma(k) = \{e \in \{1, ..., M\} \,|\, s(e) = k\} \quad \text{and} \quad \mu(k) = \{e \in \{1, ..., M\} \,|\, d(e) = k\}.$$

In the above example, we have for instance $\sigma(4) = \{2, 3, 5\}$ and $\mu(2) = \{3\}$.

One can show that $(I_e)_{e=1,...,M}$ is the unique solution to the optimization problem

$$\inf_{(I_e)_{e=1,...,M}} \quad \sum_{e=1}^{M} \frac{1}{2} R_e I_e^2, \quad \text{subject to: } J_k + \sum_{e \in \sigma(k)} I_e = \sum_{e \in \mu(k)} I_e, \quad \forall k = 1, ..., N.$$

We consider in a first step the example described above.

1. Draw the associated graph.
2. Give an interpretation of the constraints. Enumerate them all explicitely.
3. Solve the optimization problem with AMPL. *Solution:* $I = (1, 1, 0, 1, 0)$.
4. Write the KKT conditions satisfied by the solution of the problem. Give an interpretation.

We consider now a general grid.

5. Assuming that the feasible set is non-empty, prove that the problem possesses a solution.
6. Write the KKT conditions. Are they sufficient?
7. *Optional:* write an AMPL program for solving the problem associated with a general grid.

Exercice 9 (Signal processing)**.** We realize N measurements of a signal, denoted y_1,...,y_N. These measurements are supposed to be noisy and cannot be exploited as they are. We typically expect that a small portion of them is completely inaccurate. A classical denoising technique consists in solving the following optimization problem:

$$\inf_{(x_i)_{i=1,\dots,N} \in \mathbb{R}^N} \sum_{i=1}^{N} |x_i - y_i| + \alpha \sum_{i=2}^{N} (x_i - x_{i-1})^2,$$

where $\alpha > 0$ is a fixed parameter. The solution is seen as a regularized signal.

1. Explain the structure of the cost function used in the optimization problem.
2. Write an AMPL program for solving the problem. An instance of y is provided in a data file on the website of the class. Use $\alpha = 1$. You will get a more precise solution by initializing the variable x to y. Warning: the initialization of x must be done after that a numerical value has been assigned to N and y. *Expected objective value:* 15.086, obtained in 1044 iterations.

An important difficulty in the numerical resolution of the problem lies in the fact that the cost function of the problem is not differentiable, since the absolute value function is not differentiable at 0. We will see that the problem can be reformulated as a constrained problem involving smooth functions.

4. Show that for all $x \in \mathbb{R}$,

$$|x| = \min_{z \in \mathbb{R}} \ z, \quad \text{subject to: } \begin{cases} z \geq x \\ z \geq -x. \end{cases}$$

5. Reformulate the optimization problem by replacing each term of the form $|h(x)|$ by a new optimization variable z, and by adding two new constraints: $z \geq h(x)$ and $z \geq -h(x)$.
6. Write an AMPL program for solving the problem, exploiting this new formulation. *Expected objective value:* 14.394, obtained in 71 iterations.