Faculté des Sciences et Techniques 2024
Master ACSYON

Exam - Large scale optimization for machine learning

Exercice 1 — Regularization of an inverse problem.
Let $\mu > 0$, we consider the minimization problem

$$(\mathcal{P}_\mu) \qquad \min_{x\in\mathbb{R}^n} \quad \|Ax-b\|^2 + \mu\|x\|^2 \quad ,$$

and we assume that $S \neq \emptyset$, where

$$S := \{x \in \mathbb{R}^n : Ax = b\}.$$

1. Write the optimality conditions for problem $(\mathcal{P}_\mu)$ and show that its solution x_μ is unique
2. We consider now the solution x_0 to the problem

$$(\mathcal{P}_0) \quad \begin{cases} \min\limits_{x\in\mathbb{R}^n} & \|x\|^2 \\ \text{s.t.} & Ax = b \end{cases} ,$$

   show that the solution to this problem is unique.
3. Show that for all $\mu > 0$ we have $\|x_\mu\| < \|x_0\|$.
4. We assume that there is a sequence $(\mu_n)_{n\in\mathbb{N}}$ such that $\mu_n > 0$ for all n,

$$\lim_n \mu_n = 0, \qquad \text{and} \qquad \lim_n x_{\mu_n} = \tilde{x} \in \mathbb{R}^n$$

   show that $\tilde{x} \in S$.
5. Conclude that we have the following limit

$$\lim_{\mu\to 0} x_\mu = x_0.$$

Exercice 2 — Stochastic Gradient Method.
Let $n \in \mathbb{N}^*$, $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. We consider the minimization of the objective function $F : \mathbb{R}^n \to \mathbb{R}$, where for all $x \in \mathbb{R}^n$

$$F(x) = \frac{1}{2}\|Ax-b\|_2^2$$

We assume that A is symmetric and $A \succ 0$.

1. (Strong convexity). Show that there exists a constant $c > 0$ such that for all $x \in \mathbb{R}^n$, for all $y \in \mathbb{R}^n$

$$F(y) \geq F(x) + \nabla F(x)^\top (y-x) + \frac{1}{2}c\|y-x\|^2$$

2. Show that the gradient of F is Lipschitz-continuous, with Lipschitz constant $L > 0$, that is

$$\forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n, \quad \|\nabla F(y) - \nabla F(x)\|_2 \leq L\|y - x\|_2$$

3. We consider the stochastic gradient method for minimizing F at x, with the stochastic vector $g(x, K(\xi))$, where

$$g(x, k) = (a_k x - b_k)a_k^\top,$$

for $i \in \{1, \cdots, n\}$, a_i is the ith row of matrix A, b_i is the ith element of vector b and $\xi \mapsto K(\xi)$ is a random variable over the indices set $I := \{1, \cdots, n\}$.

a) We assume that K follow the uniform law of probability over the set I, compute the expectation

$$\mathbb{E}_\xi[g(x, \xi)].$$

b) Show that there exist $\mu_G \geq \mu > 0$ such that for all $x \in \mathbb{R}^n$,

$$\nabla F(x)^\top \mathbb{E}_\xi[g(x, \xi)] \geq \mu\|\nabla F(x)\|_2^2.$$

and

$$\|\mathbb{E}_\xi[g(x, \xi)]\|_2 \leq \mu_G\|\nabla F(x)\|_2$$

4. **Numerical experiments.** The file `SGmethod.py` is a Python script where the objective function is defined for $n = 30$. Both the objective value and the stochastic vector can be computed with the function `obfj_s1`.

a) Evaluate numerically the values of constant c and L.

b) Implement the stochastic gradient method with a fixed stepsize $\alpha = 0.1$. Use the initialization for x given in the Jupyter notebook. Plot the evolution of the objective value over the first 1000 iterations. What can you conclude ?

c) Explain why it is interesting to use the stochastic gradient method with diminishing stepsize in this context ?

d) Implement the stochastic gradient method with stepsize

$$\alpha_k = \frac{\beta}{\gamma + k},$$

with $\beta > \frac{1}{c\mu}$ and choose $\gamma > 0$ such that $\alpha_1 \leq \frac{\mu}{LM_G}$ with $M_G = \mu^2$.Plot the evolution of the objective value over the first 10^6 iterations.