Masters DSC/MLDM/CPS2

Optimization & Operational Research - Exam

(27/03/2018) 2h00 : personal documents allowed

Exercise 1 : Convexity and Rate of Convergence (8.5 points)

The aim of this exercise is to study the function $f_\gamma : \mathbb{R}^2 \to \mathbb{R}$ defined by :

$$f_\gamma(x,y) = \frac{1}{2}(x^2 + \gamma y^2 + 2xy) + 2x + 2y, \quad \gamma \in \mathbb{R}.$$

Part A : A study of f_γ (4.5 points)

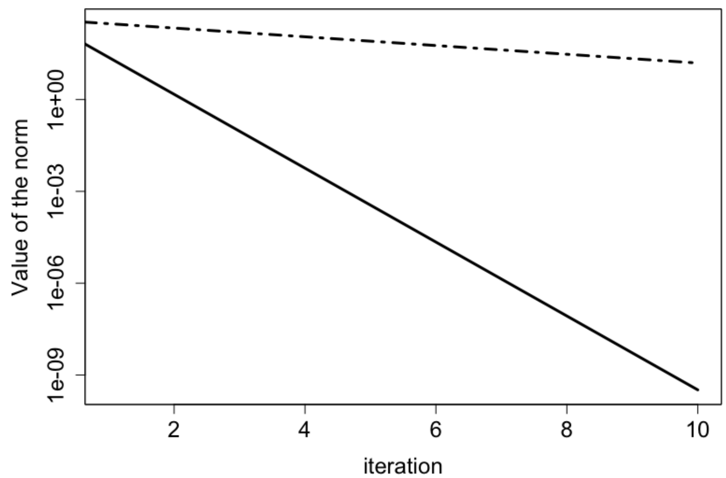
This first part is dedicated to the study of the function f_γ.

1. Study the convexity of the function f_γ.
2. Give the solution of *Euler's Equation*, i.e. the solution of the linear system $\nabla f_\gamma(x,y) = (0,0)$, for all values of γ.
3. Give the nature of the previous extrama of the function (the nature of the extremum depends on γ).
4. Show that $A = \begin{pmatrix} 1 & 1 \\ 1 & \gamma \end{pmatrix}$ and find the expression of $b \in \mathbb{R}^2$ such that, for all $u = (x\ y)^T$:
$$f_\gamma(u) = \frac{1}{2}u^T A u - b^T u,$$
5. Give the algorithm of the Gradient Descent with Optimal Step.

Part B : Rate of Convergence of the Gradient Descent with Optimal Step (4 pts)

In this part we assume that $\gamma > 1$ so that f_γ is strictly convex. The aim is to study the rate of convergence of the Gradient Descent with Optimal Step. This rate depends on the **Condition Number** of the matrix A defined by $Cond(A) = \dfrac{\lambda_{max}}{\lambda_{min}}$, where λ_{max} (resp. λ_{min}) is the largest (resp. the smallest) eigenvalue of A.

1. Compute the two eigenvalues of the matrix A.
2. Give the expression of $Cond(A)$ with respect to γ. Give an equivalent of the **Condition Number** $Cond(A)$ for large values of γ.
Hint : for large values of γ we have $(\gamma-1)^2 + 4 \simeq (\gamma-1)^2$
3. We denote by $u^\star$ the point where the function f_γ reaches its minimum and u_0 the initial point of our algorithm. The rate of convergence η of the studied algorithm is defined by $\eta = 1 - Cond(A)^{-1}$ and we have :
$$\|u_{k+1} - u^\star\|_A \leq \eta^k \|u_0 - u^\star\|_A. \tag{1}$$
The figure below illustrates the convergence of the function f_γ for two different values of γ and with the studied algorithm. We also choose $u_0 = (20\ 1)$.



Say for which curve the value of γ is the largest one. What is the impact of the Condition Number $Cond(A)$ on the rate (or speed) of convergence of the Gradient Descent according to the Inequality (1) ? Give a condition on $Cond(A)$ for which the convergence rate is fast.

4. We want to prove the Inequality (1). We denote by ρ_k the optimal learning rate at the k-*th* iteration of the algorithm.

   (a) Show that :

   $$\|u_{k+1} - u^\star\|_A^2 = \|(I - \rho_k A)(u_k - u^\star)\|_A^2.$$

   Hint : Remember that if $u^\star$ is a minimum of f_γ, then $Au^\star = b$ where A and b were defined in the previous part.

   (b) Now, we assume that for all $k \in \mathbb{N}$ we have :

   $$\|u_{k+1} - u^\star\|_A^2 \le \|I - \rho_k A\|_2^2 \|u_k - u^\star\|_A^2.$$

   Show that η^2 is an upper bound of $\|1 - \rho_k A\|_2^2$, i.e.

   $$\|I - \rho_k A\|_2^2 \le \eta^2 = \left(1 - \frac{\lambda_{min}}{\lambda_{max}}\right)^2.$$

   (c) Conclude.

Exercise 2 : (4.5 points)

Consider the following constrained optimization problem

$$\min_{x_1, x_2} \quad x_1 - x_2$$
$$\text{subject to} \quad x_1^2 + x_2^2 - 2x_2 = 0$$

1. Provide the Lagrangian formulation of this problem.
2. Deduce the Lagrange dual function associated to this problem.
3. Compute the optimum of this dual function.
4. Deduce the values that lead to an optimal solution in the primal formulation.
5. Check that the duality (weak or strong) holds. If you think you have a strong duality explain why, otherwise try to provide a justification explaining why this is not the case.