Masters DSC/MLDM/CPS2

Optimization & Operational Research - Exam

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

Correction

Part B, question 4 was considered as a bonus

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.

   In order to study the convexity of f_γ, we compute its Hessian matrix :

   $$H_{f_\gamma}(x,y) = \begin{pmatrix} 1 & 1 \\ 1 & \gamma \end{pmatrix}.$$

   The Trace of the matrix H is equal to $1+\gamma$, its Determinant is equal to $\gamma - 1$. The function f_γ is convex if both Trace and Determinant are non-negative. So f_γ is convex for $\gamma \geq 1$. It is strictly convex for $\gamma > 1$. If $\gamma < 1$, the Determinant is negative, it means that the function is neither convex or concave.

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.

   We need first to compute the Jacobian $\nabla f_\gamma(x,y)$, it is given by :

   $$\nabla f_\gamma(x,y) = \begin{pmatrix} x+y+2 & \gamma y + x + 2 \end{pmatrix}.$$

   We then have to solve the following linear system :

   $$\begin{aligned} x + y + 2 &= 0, \\ x + \gamma y + 2 &= 0. \end{aligned}$$

   By substracting the first line to the second one, we have :

   $$\begin{aligned} x &= -y - 2, \\ y(\gamma - 1) &= 0. \end{aligned}$$

   We have now two cases, depending on the value of γ :

- if $\gamma \neq 1$, then the second equation implies $y = 0$ and the first one $x = -2$.
- if $\gamma = 1$ then y can be any real value and x should be equal to $-y - 2$.

3. Give the nature of the previous extrama of the function (the nature of the extremum depends on γ).

   According to the previous questions :

   - if $\gamma > 1$, f_γ reaches its minimum at the point $(-2, 0)$ and this the global minimum because f_γ is convex.
   - if $\gamma = 1$, f_γ reaches its minimum at all the points on the line of equation $x = -y - 2$. At all this points, the function f_γ reaches its global minimum because f_γ remains convex.
   - if $\gamma < 1$, the point $(-2, 0)$ is no more minimum. This is a saddle point in this case because f_γ is neither convex or concave.

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 $u = (x\ y)^T$:

$$f_\gamma(u) = \frac{1}{2} u^T A u - b^T u,$$

   We set $b = (b_1 \quad b_2)^T$ and we develop the above expression. We have :

$$\frac{1}{2} u^T A u - b^T u \quad = \quad \frac{1}{2}\left(x^2 + \gamma y^2 + 2xy\right) - b_1 x - b_2 y$$

   By identifying the two expressions of f_γ we have :

$$b = (-2 \quad -2)^T.$$

5. Give the algorithm of the Gradient Descent with Optimal Step.

   I give the algorithm in the context of the problem. It was also possible to give it for a general function f.

   1) Choose an initial point u_0 in the domain of definition of the function f_γ
   2) Repeat for all $k \in \mathbb{N}$
      - Compute $\nabla f_\gamma(u_k) = Au_k - b$
      - Choose the optimal learning rate ρ such that :

$$\rho_k = \underset{\rho \in \mathbb{R}_+}{Argmin}\ f(u_k - \rho_k \nabla f_\gamma(u_k)).$$

        So that ρ_k is equal to $\dfrac{\|Au_k - b\|_2^2}{\|Au_k - b\|_A^2}$ in this context.
      - Update : $u_{k+1} = u_k - \rho \nabla f_\gamma(u_k)$.
   3) Till $\|\nabla f(u_{k+1})\|_2 \leq \varepsilon$

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.

   The eigenvalues are the roots of the polynom :

$$det\,(A - \lambda I) = \begin{pmatrix} 1-\lambda & 1 \\ 1 & \gamma - \lambda \end{pmatrix} = (1-\lambda)(\gamma-\lambda) - 1 = \lambda^2 - \lambda(\gamma+1) + \gamma - 1.$$

   The roots are defined by :

$$\lambda_\pm = \frac{\gamma + 1 \pm \sqrt{\Delta}}{2},$$

   where $\Delta = (\gamma+1)^2 - 4(\gamma-1) = \gamma^2 - 2\gamma + 5 = (\gamma-1)^2 + 4$.

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* $(\gamma-1)^2 + 4 \simeq (\gamma-1)^2$

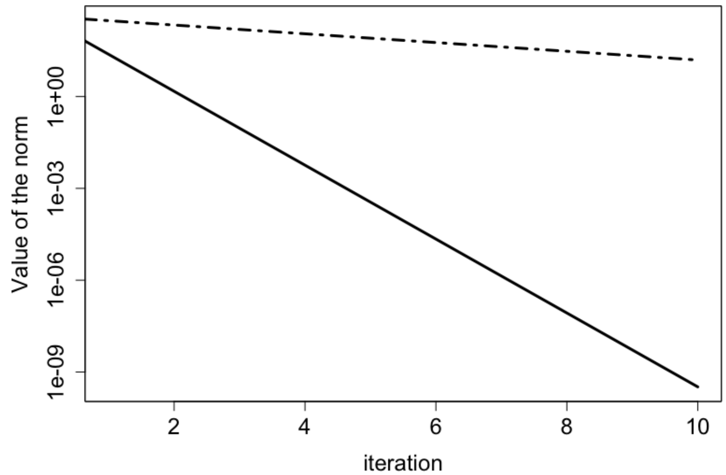
   For large values of γ we have $\lambda_\pm \simeq \dfrac{\gamma + 1 \pm (\gamma - 1)}{2}$. So that the Condition Number $Cond(A)$ can be approximated by :

$$Cond(A) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{\lambda_+}{\lambda_-} = \frac{\frac{\gamma+1+\gamma-1}{2}}{\frac{\gamma+1-\gamma+1}{2}} \simeq \frac{\gamma}{1} = \gamma$$

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 :

$$\|u_{k+1} - u^\star\|_A \leq \eta^k \|u_0 - u^\star\|_A. \quad (1)$$

   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.

According to Inequality (1), the convergence will be faster if η is close to 0. By definition of η it means that $Cond(A)^{-1}$ should be close to 1.
We have seen that $Cond(A) = \gamma$ so $Cond(A)^{-1} = \dfrac{1}{\gamma}$.
So the larger the value of γ is the slower the convergence is and conversly.
The Gradient Descent with optimal step converges rapidly toward $(-2, 0)$ if γ is close to 1. So the dashed line represents the case where the value of γ is the largest one.

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.

   It is enough to show that the two vector in the norm are the same. We develop the right hand side of the equation.

   $$\begin{aligned}(I - \rho_k A)(u_k - u^\star) &= u_k - u^\star - \rho_k A u_k - \rho_k A u^\star,\\ &= u_k - u^\star - \rho_k A u_k - \rho_k b,\\ &= u_k - \rho_k(Au_k - b) - u^\star,\\ &= u_{k+1} - u^\star.\end{aligned}$$

   The second line uses the fact that $Au^\star = b$.

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

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

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

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

You have to give an upper bound $\|I - \rho_k A\|_2^2$. First, you shall remember that the optimal learning rate is given by :

$$\rho_k = \frac{\|Au_k - b\|_2^2}{\|Au_k - b\|_A^2}.$$

Furthermore, for all PSD matrices A.

$$\lambda_{min}(A)I \leqslant A \leqslant \lambda_{max}(A)I,$$

where the inequalities mean that the eigenvalues of the matrix on the left-handside are less than all the eigenvalues in the middle and so on.
So, by multiplying on the left by u^T and by u on the right for any vector u, we get :

$$\lambda_{min}(A)\|u\|_2 \leq \|u\|_A \leq \lambda_{max}(A)\|u\|_2.$$

Now, to upper bound $\|I - \rho_k A\|_2^2$, we need to give a lower bound on both A and ρ_k (because of minus sign) :

- for A we have the following lower bound $\lambda_{min}(A)I \leqslant A$
- $\rho_k \geq \dfrac{\|Au_k - b\|_2^2}{\lambda_{max}(A)\|Au_k - b\|_2^2} = \dfrac{1}{\lambda_{max}(A)}$.

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

(c) Conclude.

We conclude using a chain rule :

$$\begin{aligned}
\|u_k - u^\star\|_A &\leq \eta\|u_{k-1} - u^\star\|_A, \\
&\leq \eta^2\|u_{k-2} - u^\star\|_A, \\
&\leq \eta^3\|u_{k-3} - u^\star\|_A, \\
&\leq ..., \\
&\leq \eta^k\|u_0 - u^\star\|_A.
\end{aligned}$$

Exercise 2 : (4.5 points)

Consider the following constrained optimization problem

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

1. Try to represent draw the set of constraints and the function in the $(x_1, x_2)-$space and try to see the solution of this minimization problem.

2. Provide the Lagrangian formulation of this problem.
3. Deduce the Lagrange dual function associated to this problem.
4. Compute the optimum of this dual function.
5. Deduce the values that lead to an optimal solution in the primal formulation.
6. 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.