École d'ingénieurs
Télécom Physique
Université de Strasbourg

Information Sciences and Technology for Healthcare | HealthTech

The Interdisciplinary thematic institutes of the University of Strasbourg & cnrs & Inserm funded under the Excellence Initiative program

REPORT LAB SESSION

INTRODUCTION TO OPTIMIZATION

Jesse de Oliveira Santana Alves
Luis Villamarin

Professor: Omran Hassan

April 2023

Part 1.

For the first part of the lab session, there is a file named *main_part1.m*, where all the functions implemented can be tested.

Question 1.1

The first linear programing method (file *Exhaustive_LP.m*) used the following values:

- $A = \begin{bmatrix} 1 & 5 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}$,
- $b = [40\ 20\ 12]^T$,
- $c = [-3\ -5\ \ 0\ \ 0\ \ 0]^T$.

And after 10 iterations, i.e., testing 10 different bases, it was obtained the following results.

- Cost function: $f = -50$,
- Optimal BFS: $x = [5\ 7\ 0\ 3\ 0]^T$,
- Basis: $B = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

This method is relatively easy to implement, however, it is necessary to test each base to find the optimal solution. And this, can be an issue for big linear systems. This method can be used to solve linear programming in standard form. Therefore, it is known that any linear programming problem involving inequalities can be converted to standard form, that is, a problem involving linear equations with nonnegative variables (*Reference for Linear Programming book*):

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $m < n$, $rank\boldsymbol{A} = m$ and $b \geq 0$.

Question 1.2

To the second method implemented (file *LP_Simplex.m*) the algorithm was tested with the same values of matrix A and vectors c and b. And after 3 iterations, it was obtained the following results.

- Cost function: $f = -50$,
- Optimal BFS: $x = [5\ 7\ 0\ 3\ 0]^T$,
- Basis: $B = [2\ 4\ 1]^T$.

Where the Basis B represents the index number for each vector a_i. This method was more difficult to implement than the first one, however, proved to be faster to find the minimizer vector with less iteration numbers, at least for the systems used in this lab.

Question 1.3

For the Two-Phase Simplex Method, given the following input arguments:

$$A = \begin{bmatrix} 4 & 2 & -1 & 0 \\ 1 & 4 & 0 & -1 \end{bmatrix}$$

$$b = [12 \quad 6]$$

$$c = [2\ \ 3\ \ 0\ \ 0\ \ 0]^T$$

the results are:

$$f = -7.7143$$

$$B = [1 \quad 2]$$

$$x = \begin{bmatrix} 2.5714 \\ 0.8571 \\ 0 \\ 0 \end{bmatrix}$$

Part 2.

Question 2.1

A Golden Section algorithm was written as a method of optimization for the function:

$$\boldsymbol{f(x)} = x^4 + 4x^3 + 9x^2 + 6x + 6$$

Where ρ is the Golden Ratio constant that determines the size of the search interval at each iteration of the algorithm, given as result the following values:

$$\boldsymbol{xmin} = -0.4473$$

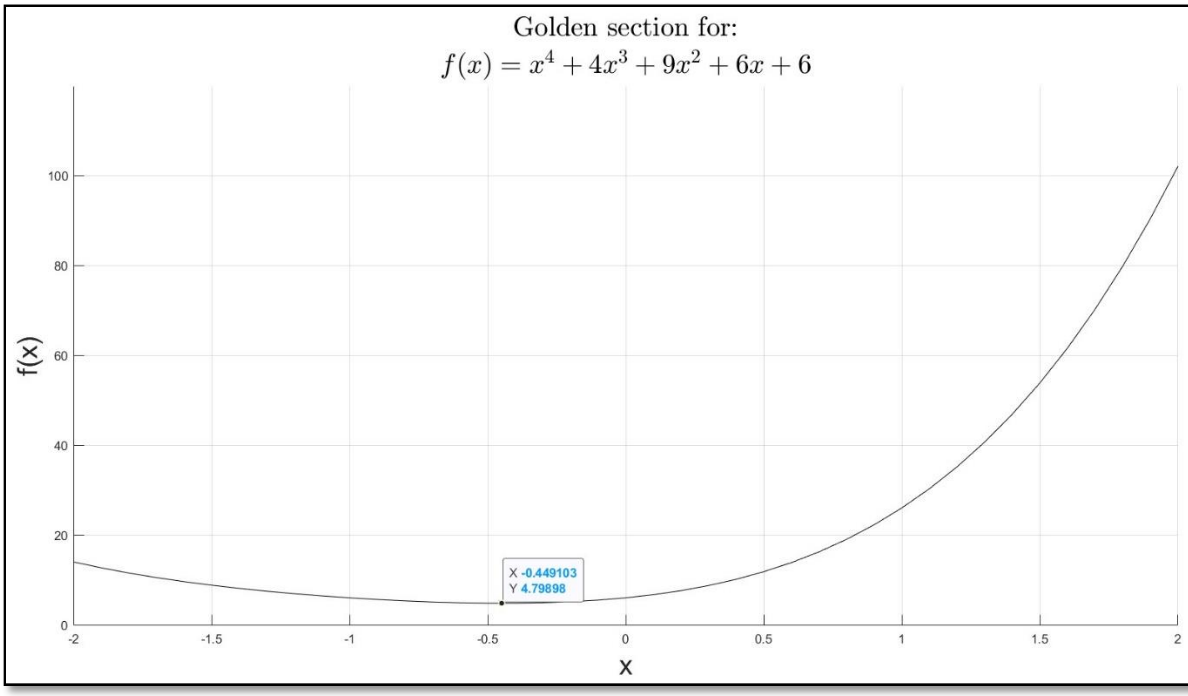

Figure 1 – Minimizer and Minimum cost using Golden Section.

Question 2.2

The Matlab function called *Newton_Secant.m* shows the following results.

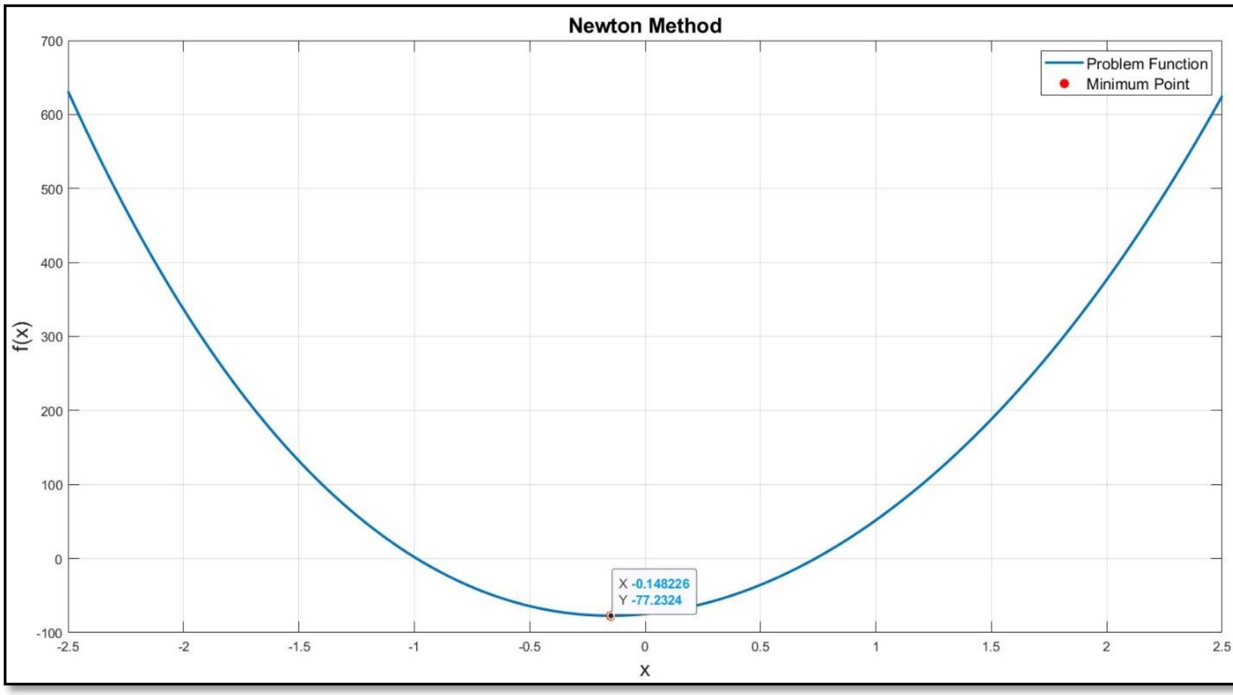

Figure 2 – Minimizer and Minimum cost using Newton Method.

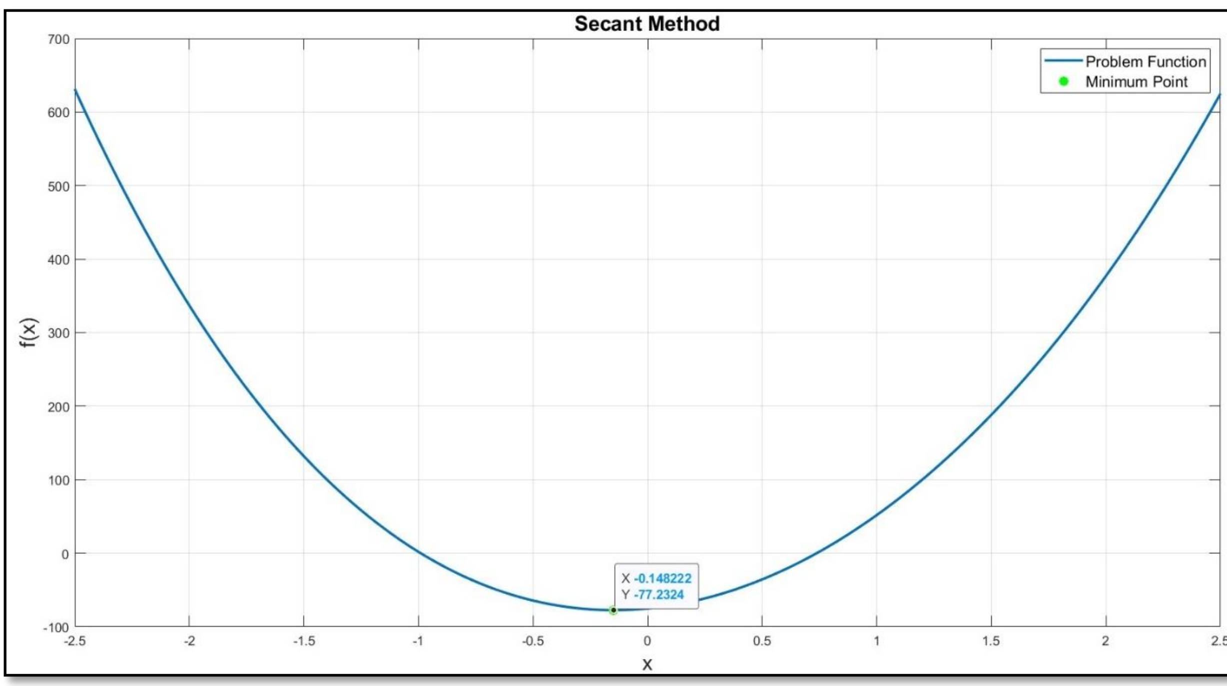

Figure 3 – Minimizer and Minimum cost using Secant Method.

As both methods used the same function, stop criterion and the same initial point, it is possible to compare the number of iterations and accuracy for each one. The accuracy here is measured comparing the value of cost function of each method with the "true cost function", obtained by using the function *min ()* with a vector *f(x)* with high precision.

The Newton Method reaches the minimizer with 3 iterations and accuracy of 1.65×10^{-9}. On the other hand, the Secant Method reach the minimizer with 4 iterations and a better accuracy of 2.70×10^{-13}.

Part 3.

Question 3.1

1. The first question is demonstrated in the Annex A.

2. For this implementation, is expected that the minimizers of the Rosenbrock's function are $[x1\ x2] = [1\ 1]$, as it was proofed in the first item. However, using the Steepest Descent Method with Newton Method in this question, for the stopping criterion given, the algorithm did not find the minimum. Indeed, the value found was $[x1\ x2] = [0.49\ \ 0.23]$, thus proving that the Steepest Descent does not have a good performance for the Rosenbrock's function, where there exists a zig zag convergence near the optimal point. After analyzing the level set of the function (figure below) and evaluate that the function is convex at x2 but not convex at x1, it is possible to conclude that the absence of convexity promotes this zig zag effect in the algorithm, leading to the need for many the function evaluations to converge.

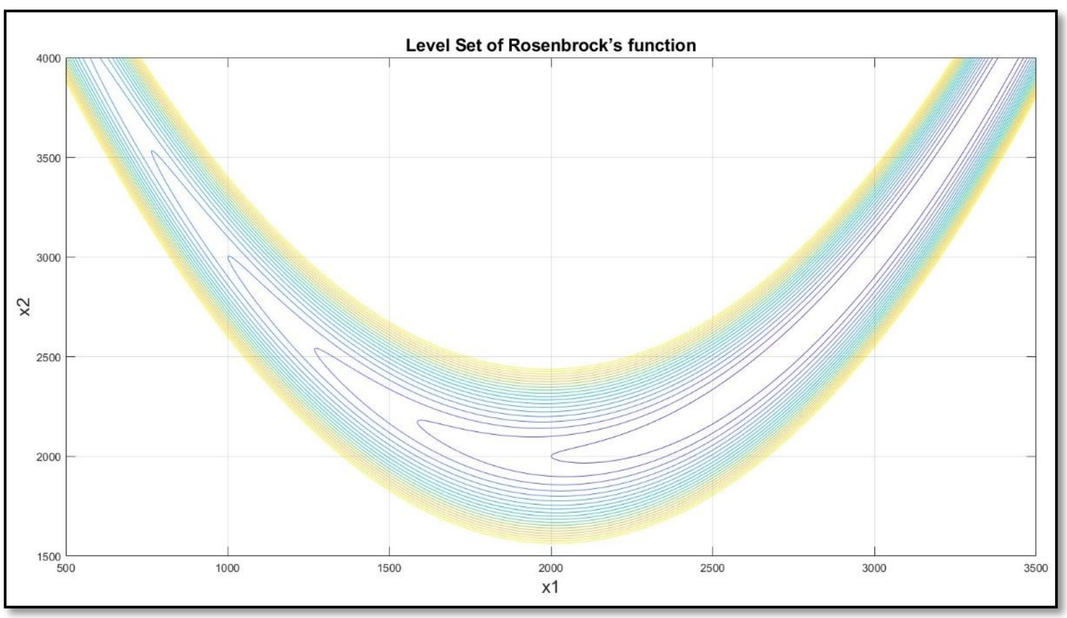

Figure 4 – Rank for the three cases.

Question 3.2

The question 3.2 is also realized in the Annex B.

Question 3.3

1. To categorize the speed of convergence of each of the cases presented, it is possible to evaluate the lambda coefficient of each one of them and through a relationship between the maximum and the minimum, rank them as follows:

$$Case\ 1) \quad \frac{\lambda \max(Q)}{\lambda min(Q)} = 1 \qquad Fastest$$

$$Case\ 3) \quad \frac{\lambda \max(Q)}{\lambda min(Q)} = \frac{2}{1} = 2 \qquad Second\ Fastest$$

$$Case\ 2) \quad \frac{\lambda \max(Q)}{\lambda min(Q)} = \frac{10}{1} = 10 \qquad Slowest$$

The computation to rank all the three cases in term of speed is also done in the file *Steepest_descent_question3_3.m*, in the "RANK THE METHODS" part. There, is demonstrated that:

```
=========== Clearly the rank, from faster to slower, is : ===========
Rank 1 = Case 1
Rank 2 = Case 3
Rank 3 = Case 2
```

Figure 5 – Rank for the three cases.

2. For the routine implementation, also in the *Steepest_descent_question3_3.m* file, is demonstrated that the plots of the iterations (figures below) match, in terms of speed of convergence, with the evaluation of the condition number of Q.

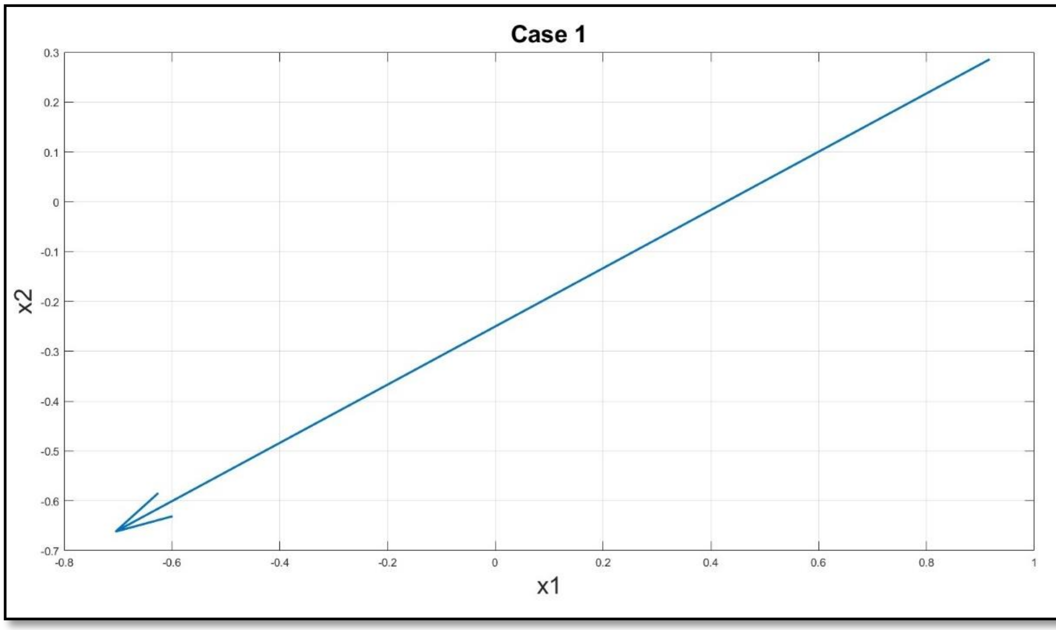

Figure 6 – Iteration plot Case 1.

It is possible to see that, since the matrix Q has the same lambda values in the main diagonal, the condition number is always one and the algorithm converges in the first iteration.

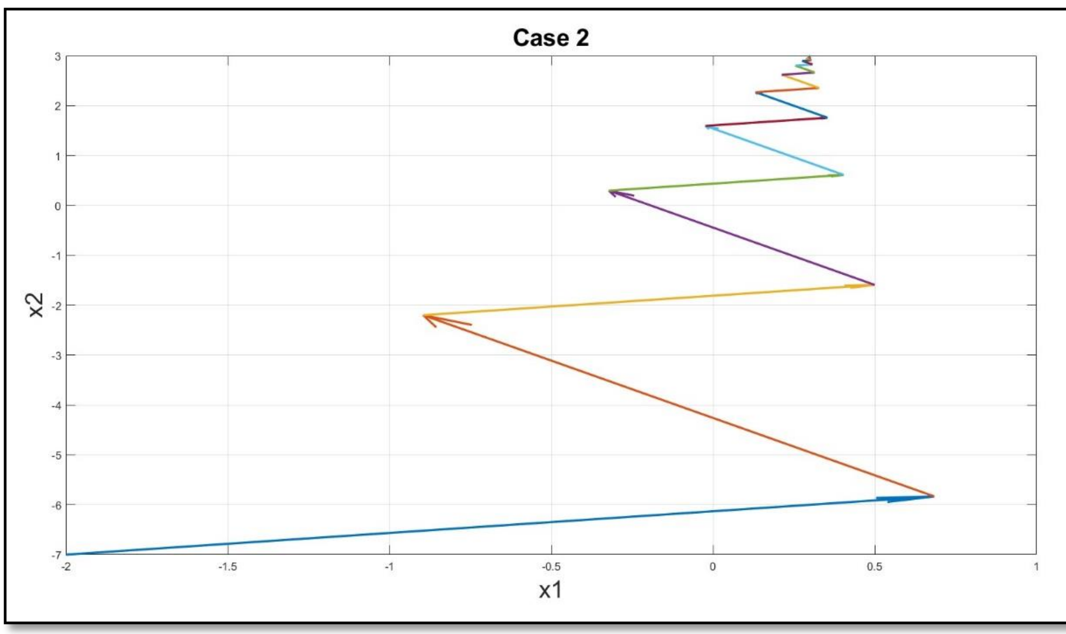

Figure 7 – Iteration plot Case 2.

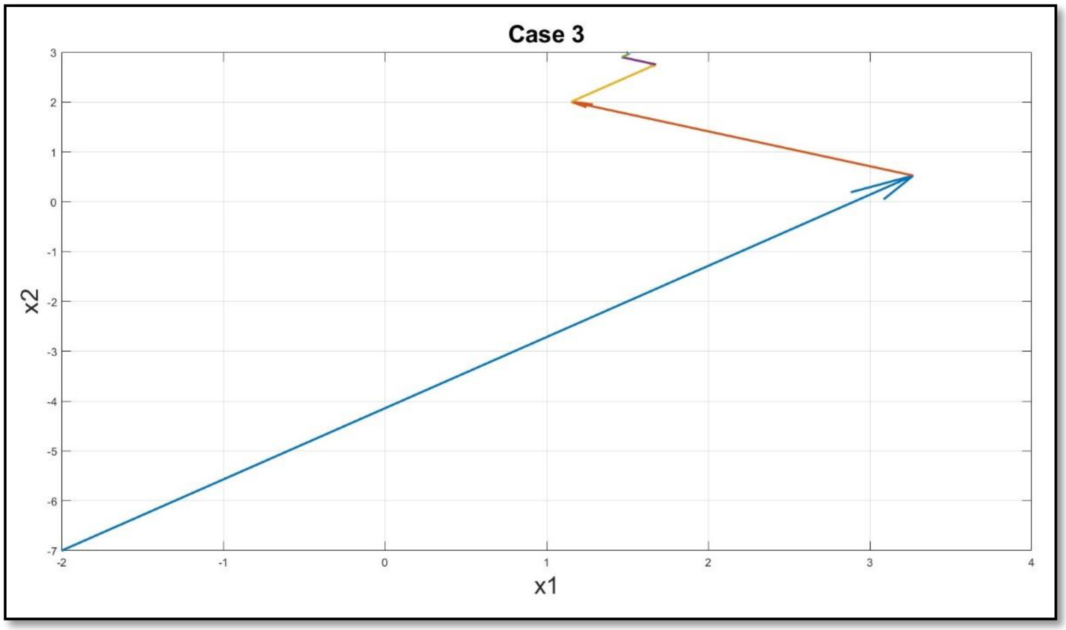

Figure 8 – Iteration plot Case 3.

Annex A - Question 3.1

vendredi 31 mars 2023 17:16

3.1 Consider the problem of finding the minimizer of Rosenbrock's function :

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) = 100(x_2 - x_1{}^2)^2 + (1 - x_1)^2$$
$$\text{s.t.} \quad \mathbf{x} \in \mathbb{R}^2$$

1. Show analytically that $[1 \;\; 1]^T$ is the unique global minimizer.

Based on the following Theorem

Theorem 5 (Second-Order Sufficient Condition)

*Consider $\mathcal{S}$ a subset of $\mathbb{R}^n$ and a twice continuously differentiable function $f : \mathcal{S} \to \mathbb{R}$ and an **interior point** $\mathbf{x}^\star$. If*

- $\nabla f(\mathbf{x}^\star) = \mathbf{0}$
- $D^2 f(\mathbf{x}^\star) > 0$ *(the Hessian matrix at $\mathbf{x}^\star$ is positive definite)*

then $\mathbf{x}^\star$ is a strict local minimizer of $f(\cdot)$.

1

$$\nabla f = [f_{x_1} \;\; f_{x_2}]^T$$

$$f_{x_1} = \frac{\partial f}{\partial x_1} = 200(x_2 - x_1^2)(-2x_1) + 2(1 - x_1)(-1)$$

$$\boxed{f_{x_1} = -400\,x_1(x_2 - x_1^2) + 2x_1 - 2} \quad (\mathrm{I})$$

$$f_{x_2} = \frac{\partial f}{\partial x_2} \rightarrow \boxed{f_{x_2} = 200(x_2 - x_1^2)} \quad (\mathrm{II})$$

Then, to reach $\nabla f = 0$:

$$f_{x_2} = 0 \Rightarrow 200(x_2 - x_1^2) = 0 \Rightarrow \boxed{x_2 = x_1^2} \quad (\mathrm{III})$$

$(\mathrm{III}) \rightarrow (\mathrm{I})$ and $f_{x_1} = 0$:

$$-400\,x_1\cancel{(x_1^2 - x_1^2)}^{\to 0} + 2x_1 - 2 = 0 \Rightarrow \boxed{x_1 = 1} \Rightarrow \boxed{x_2 = 1}$$

Then, the unique global minimizer is $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [1 \quad 1]^T$.

Also, testing the point [1 1] in the second SONC condition:

2 $D^2 f(x^*) = \begin{bmatrix} f_{x_1 x_1} & f_{x_1 x_2} \\ f_{x_2 x_1} & f_{x_2 x_2} \end{bmatrix}$, f is twice diff, so $f_{x_1 x_2} = f_{x_2 x_1}$.

$$\begin{cases} f_{x_1 x_1} = -400\left[x_2 - 3x_1^2\right] + 2 \\ f_{x_1 x_2} = f_{x_2 x_1} = -400 x_1 \\ f_{x_2 x_2} = 200 \end{cases}$$

$$D^2 f(x)\Big|_{\substack{x_1=1 \\ x_2=1}} = \begin{bmatrix} -400(-2)+2 & -400 \\ -400 & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

Then, based on:

Theorem 9

A symmetric matrix $M = M^T \in \mathbb{R}^{n \times n}$ is positive definite (or positive semidefinite) if and only if all eigenvalues of M are positive (or nonnegative).

$$P(\lambda) = (802-\lambda)(200-\lambda) - 400^2 = 160.400 - 802\lambda - 200\lambda + \lambda^2 - 400^2$$

$$P(\lambda) = \lambda^2 - 1002\lambda + 400 = 0 \Rightarrow$$

$\boxed{\lambda_1 = 0{,}39 \quad \text{and} \quad \lambda_2 = 1001{,}6}$ $\Rightarrow D^2 f(x)\Big|_{\substack{x_1=1 \\ x_2=1}}$ is positive definite ✓

Then, $[x_1 \ x_2] = [1 \ 1]$ is a strict local minimizer.

Annex B - Question 3.2

vendredi 31 mars 2023 17:16

3.2 Consider the following problem :

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = (2{x_1}^2 - x_2)^2 + 3{x_1}^2 - x_2 \\ \text{s.t.} \quad & \mathbf{x} \in \mathbb{R}^2 \end{aligned}$$

1. Perform theoretically one iteration of the steepest descent method (using an exact line) starting from $x_0 = [1/2 \;\; 5/4]^T$.
2. Is the function convex at x_1 ?

1 Computing the gradient of $f(x)$:

$$\nabla f = [f_{x_1} \;\; f_{x_2}]^T$$

$$f_{x_1} = \frac{\partial f}{\partial x_1} = 2(2x_1^2 - x_2)(4x_1) + 6x_1 = 8x_1(2x_1^2 - x_2) + 6x_1$$

$$f_{x_2} = \frac{\partial f}{\partial x_2} = 2(2x_1^2 - x_2)(-1) - 1 = -4x_1^2 + 2x_2 - 1$$

$$\Rightarrow \boxed{\nabla f(x_1, x_2) = \begin{bmatrix} 8x_1(2x_1^2 - x_2) + 6x_1 \\ -4x_1^2 + 2x_2 - 1 \end{bmatrix}}$$

Considering $x_0 = [\frac{1}{2} \;\; \frac{5}{4}]^T$ yields

$$x_{K+1} = x_K - \alpha_K \nabla f(x_K)$$

for $K = 0$,

$$x_1 = x_0 - \alpha_K \cdot \nabla f(x_0)$$

$$x_1 = \begin{bmatrix} \frac{1}{2} & \frac{5}{4} \end{bmatrix}^T - \alpha_K \begin{bmatrix} 8 \cdot 0{,}5\,(2 \cdot (0{,}5)^2 - 5/4) + 6(0{,}5) \\ -4(0{,}5)^2 + 2 \cdot \frac{5}{4} - 1 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} 0{,}5 \\ 1{,}25 \end{bmatrix} - \alpha_K \begin{bmatrix} 4(0{,}5 - 1{,}25) + 3 \\ -1 + 2{,}5 - 1 \end{bmatrix} = \begin{bmatrix} 0{,}5 \\ 1{,}25 \end{bmatrix} - \alpha_K \begin{bmatrix} 0 \\ 0{,}5 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} 0{,}5 \\ 1{,}25 - 0{,}5\,\alpha_K \end{bmatrix}$$

Considering $\alpha_K = 0{,}1$, for instance, it yields: $x_1 = \begin{bmatrix} 0{,}5 \\ 1{,}2 \end{bmatrix}$

Finnaly the stopping criteria is evaluated, for example:

$$\frac{\| x_{K+1} - x_K \|}{\| x_K \|} < \varepsilon_1 \quad \Rightarrow \quad \frac{\| x_1 - x_0 \|}{\| x_0 \|} = 0{,}0371 < \varepsilon_1$$

If $\frac{\| x_{K+1} - x_K \|}{\| x_K \|} < \varepsilon_1$, you found your minimizer.

Else, you update $x_K = x_{K+1}$ and compute again!

$$x_{K+1} = x_K - \alpha_K \nabla f(x_K)$$

2

If the function is twice differentiable, it is enough to check that the second derivative is non negative. Based on that:

Computing the second derivative for x_1:

$$f_{x_1 x_1} = 48x_1^2 - 8x_2 + 6$$

Search for non negative values:

$$f_{x_1 x_1} < 0 \Rightarrow 48x_1^2 - 8x_2 + 6 < 0 \Rightarrow 48x_1^2 + 6 < 8x_2$$

$$\boxed{x_2 > 6x_1^2 + 0{,}75} \quad (*)$$

It means that, every time $(*)$ is met, the function $f_{x_1 x_2} < 0$, therefore the function $f(x_1, x_2)$ is not convex at x_1.