Optimization Methods **Spring 2024**

Student: Birindelli Leonardo **Discussed with:**

Fourth assignment

Due date: Monday, 3 June 2024, 11:59 PM

Exercise 1 (20/100)

Consider the quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as:

$$f(x) = 7x^2 + 4xy + y^2 \tag{1}$$

where $\mathbf{x} = (x, y)^T$.

1. Write this function in canonical form, i.e. $f(x) = \frac{1}{2}x^T Ax - b^T x + c$, where A is a symmetric matrix.

   The function $f(x)$ can be written according to the canonical form definition previously described as follows:

$$f(x) = \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 14 & 4 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

   It can be easily noticed that the matrix A is symmetric, the vector b is equal to $\begin{bmatrix} 0 & 0 \end{bmatrix}$ because there are not any 1° grade polynomials and the constant c is equal to 0. It is possible to prove that the canonical form is correct for the function $f(x)$ by computing the product of the matrices and the vectors as follows:

$$\begin{aligned} f(x) &= \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 14 & 4 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + 0 \\ &= \frac{1}{2}(14x^2 + \underline{4xy} + \underline{4xy} + 2y^2) \\ &= 7x^2 + 4xy + y^2 \end{aligned}$$

2. Describe briefly how the Conjugate Gradient (CG) Method works and discuss whether it is suitable to minimize f from equation (1). Explain your reasoning in detail **(max. 30 lines)**.

   The Conjugate Gradient method is an optimization algorithm for solving linear systems of equations in the form $Ax = b$ where A is a symmetric positive definite matrix. The linear system can be interpreted as a minimization problem for a quadratic function with the matrix A s.p.d. which it can be described as :

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2}x^T Ax - b^T x \tag{2}$$

The CG method is based on the idea of finding the minimum of the function $f(x)$ by generating a set of vectors that guarantee the conjugacy property with respect to the s.p.d matrix A in a lighter way, this property guarantees the method reaches the convergence in at most n steps. In this particular case, the iterative method is suitable to minimize the function $f(x)$ from equation (1) because previously I proved that the function could be written in the canonical form and it was possible to notice that the one correspond to the exact quadratic form expression as described in the minimization problem (2) in this case, thus,we know the matrix A is symmetric positive definite, which is the necessary conditions for the method to work properly. Note that the CG is used generally for problems in which we have a large sparse matrix due to efficiency in terms of memory usage and computational complexity at each iteration (that requires a matrix-vector multiplication) and its negative point is the fact that it is affected by rounding errors.In this case where the matrix A is very small, it would have been better to use the Gauss elimination or a factorization methods instead of the CG.

Exercise 2

Consider the following constrained minimization problem for $\mathbf{x} = (x, y, z)^T$

$$\begin{aligned} &\min f(x) := -3x^2 + y^2 + 2z^2 + 2(x + y + z) \\ &\text{s.t. } c(\mathbf{x}) = x^2 + y^2 + z^2 - 1 = 0 \end{aligned} \tag{3}$$

Write down the Lagrangian function and derive the KKT conditions for 3

To begin with, the Lagrangian function is a mathematical tool used to solve minimization or maximization optimization problems in which the objective function is subjected to a set of constraints that describe which is the feasible area of solutions. The Lagrangian function is defined as follows:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{n} \lambda_i c_i(x)$$

where $x \in \mathbb{R}^n$, $f(x)$ is the objective function, $c_i(x)$ are the constraints and λ_i are the Lagrange multipliers which are scalar values used as coefficients in the linear combination of the constraint gradients which allow to satisfy those conditions while we optimize the objective function. They are required to guarantee that the given solution is both a stationary point of the initial function and stationary point for the Lagrangian function.
As it is possible to notice, the optimization problem is subjected to only one equality constraint in this problem, thus, it can be possible to semplify the Lagrangian function (3) as follow:

$$\mathcal{L}(x, \lambda) = f(x) - \lambda c(x) \tag{4}$$
$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) - \lambda \nabla c(x)$$

As we proceed, it becomes important to also introduce the *Karush-Kuhn-Tucker conditions*: those requisites, also known as KKT conditions, are a set of first-order necessary conditions used to define if a solution is optimal for a constrained optimization problem. The KKT conditions are a generalization of the Lagrange multipliers method to equality and inequality constraints.
*The theorem behind those necessary conditions, as introduced at the section **"Theorem 12.1 (First-Order Necessary Conditions)"** page 321 of "Jorge Nocedal and Stephen J Wright. Numerical optimization. Springer, 1999"(1)* describes that:

Suppose that x^* is a local solution of (12.1), that the functions f and c_i in (12.1) are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, *with components* λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, *such that the following conditions are satisfied at* (x^*, λ^*)

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\
c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, \quad \text{for all } i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, \quad \text{for all } i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.
\end{aligned}$$

where $\mathcal{E}$ and $\mathcal{I}$ are the sets of equality and inequality constraints, respectively, and the LICQ refers to the *Linear Independence Constraint Qualification property* that must be verified by the active constraint gradients at the point x^*. This attribute is important to guarantee that the point x^* is a local solution for the problem by verifying that the gradients of constraint are not linearly dependent between one another at that specific point; a deeper explanation of the LICQ property and the active set can be found respectively in the sections **"Definition 12.4 (LICQ)"** *at page 320* and **"Definition 12.1"** *at page 308 of "Jorge Nocedal and Stephen J Wright. Numerical optimization. Springer, 1999"*(1).

In the case of the problem 3, the KKT conditions can be summarized in the following conditions :

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\
c_0(x^*) &= 0, \quad \text{for all } i \in \mathcal{E}, \\
\lambda_0^* c_0(x^*) &= 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}
\end{aligned}$$

where $\mathcal{I} = \emptyset$, $\mathcal{E} = \{0\}$ in which $i = 0$ correspond to the unique equality constraint in the problem.

Exercise 3

1. Read the chapter on Simplex method, in particular the section 13.3 The Simplex Method, in Numerical Optimization, Nocedal and Wright. Explain how the method works, with a particular attention to the search direction.

2. Consider the following contrained minimization problem, $\mathbf{x} = (x_1, x_2)^T$;

$$\min_x f(x) = 4x_1 + 3x_2 \tag{5}$$

subject to:

$$\begin{aligned}
6 - 2x_1 - 3x_2 &\geq 0 \\
3 + 3x_1 - 2x_2 &\geq 0 \\
5 - 2x_2 &\geq 0 \\
4 - 2x_1 - x_2 &\geq 0 \\
x_2 &\geq 0 \\
x_1 &\geq 0
\end{aligned}$$

a) Sketch the feasible region for this problem.

Firstly, let's analyze the inequality constraints to determine the boundaries of the feasible region. The optmization problem is restricted by the following constraints:

$$
\begin{aligned}
6 - 2x_1 - 3x_2 &\geq 0 \\
3 + 3x_1 - 2x_2 &\geq 0 \\
5 - 2x_2 &\geq 0 \\
4 - 2x_1 - x_2 &\geq 0 \\
x_2 &\geq 0 \\
x_1 &\geq 0
\end{aligned}
$$

Now, by solving the system of inequality constraints in terms of x_2, we can determine the boundaries of the feasible region of the problem (5), what I obtained is the following:

$$
\begin{cases}
6 - 2x_1 - 3x_2 \geq 0 & \rightarrow x_2 \leq -\frac{2x_1}{3} + 2 \\
3 + 3x_1 - 2x_2 \geq 0 & \rightarrow x_2 \leq \frac{3}{2}x_1 + \frac{3}{2} \\
5 - 2x_2 \geq 0 & \rightarrow x_2 \leq \frac{5}{2} \\
4 - 2x_1 - x_2 \geq 0 & \rightarrow x_2 \leq -2x_1 + 4 \\
x_1 \geq 0 \\
x_2 \geq 0
\end{cases}
$$

Secondly , I represent the feasible area marked by the constraints on the Cartesian plane and I obtain:

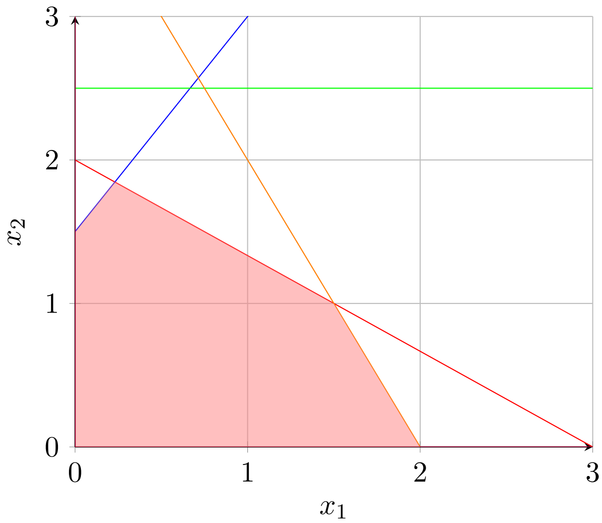

Figure 1: Feasible region for the problem (5)

Note that I decide to color each straight line with different dyes in order to easily understand the plane: The red line represents the line equation of the constraint $x_2 \leq -\frac{2x_1}{3} + 2$, the green line is the equation corresponding to the constraint $x_2 \leq \frac{5}{2}$, the blue one is referring to the line equation of the constraint $x_2 \leq \frac{3}{2}x_1 + \frac{3}{2}$, the orange one is referring to the line equation of the constraint $x_2 \leq -2x_1 + 4$ and ,in conclusion, the barely visible purple lines correspond to the non-negativity constraints $x, y \geq 0$

//TODO: to check the correctness of the answer and the explanation of the feasible region

b) Which are the basic feasible points of the problem (5)? Compute them by hand using the geometrical interpretation and find the optimal point $\mathbf{x}^*$ that minimizes f subject to the constraints. First of all , to find the basic feasible points, I need to find the intersection points between the constraints. The basic feasible points are the vertices of the feasible region.

As the theorem described in the section **"Theorem 13.3 "** *page 365 of "Jorge Nocedal and Stephen J Wright. Numerical optimization. Springer, 1999"*(1) affirms that all the basic feasible points are always located at vertexes of the feasible region. Therefore, it is required to look at the point coordinates of the ones that delimitate the feasible area by computing the intersections between the different inequalities.

To begin with, it is really easy to notice that one point of the basic feasible points is the origin $A(0,0)$ because it is the intersection between the non-negativity constraints $x_1 \geq 0$ and $x_2 \geq 0$.
Subsequently, I aimed to find the intersection point between y-axis $x_1 = 0$ and $x_2 \leq \frac{3}{2}x_1 + \frac{3}{2}$, thus, what I obtained is the following result:

$$\begin{cases} x_1 = 0 \\ x_2 = \frac{3}{2}x_1 + \frac{3}{2} \end{cases} \rightarrow \begin{cases} x_1 = 0 \\ x_2 = \frac{3}{2} \cdot 0 + \frac{3}{2} = \frac{3}{2} \end{cases}$$

The intersection point between those two lines is $B(0, \frac{3}{2})$.
Secondly, I found the intersection point between $x_2 \leq \frac{3}{2}x_1 + \frac{3}{2}$ and $x_2 \leq -\frac{2x_1}{3} + 2$, thus, the computations are the following:

$$\begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ x_2 = \frac{3}{2}x_1 + \frac{3}{2} \end{cases} \rightarrow \begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ -\frac{2x_1}{3} + 2 = \frac{3}{2}x_1 + \frac{3}{2} \end{cases}$$

$$\begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ \frac{-4x_1+12}{\cancel{6}} \cdot \cancel{6} = \frac{9x_1+9}{\cancel{6}} \cdot \cancel{6} \end{cases} \rightarrow \begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ -4x_1 - 9x_1 = -12 + 9 \end{cases}$$

$$\begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ -13x_1 = -3 \end{cases} \rightarrow \begin{cases} x_2 = -\frac{2 \cdot \frac{3}{13}}{3} + 2 \\ x_1 = \frac{3}{13} \end{cases}$$

$$\begin{cases} x_2 = -\frac{2}{13} + 2 \\ x_1 = -\frac{3}{13} \end{cases} \rightarrow \begin{cases} x_2 = \frac{24}{13} \\ x_1 = \frac{3}{13} \end{cases}$$

The intersection point between those two lines is $C(\frac{3}{13}, \frac{24}{13})$.
Thirdly, I found the intersection point between $x_2 \leq -\frac{2x_1}{3} + 2$ and $x_2 \leq -2x_1 + 4$, thus, the computations are the following:

$$\begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ x_2 = -2x_1 + 4 \end{cases} \rightarrow \begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ -\frac{2x_1}{3} + 2 = -2x_1 + 4 \end{cases}$$

$$\begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ \frac{-2x_1+6}{\cancel{3}} \cdot \cancel{3} = -\frac{-6x_1+12}{\cancel{3}} \cdot \cancel{3} \end{cases} \rightarrow \begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ -2x_1 + 6 = -6x_1 + 12 \end{cases}$$

$$\begin{cases} x_2 = -\frac{2x_1}{3} + 2 \\ 4x_1 = 6 \end{cases} \rightarrow \begin{cases} x_2 = -\frac{2 \cdot \frac{3}{2}}{3} + 2 \\ x_1 = \frac{\cancel{6}^3}{\cancel{4}^2} \end{cases}$$

$$\begin{cases} x_2 = -1 + 2 \\ x_1 = \frac{3}{2} \end{cases} \rightarrow \begin{cases} x_2 = 1 \\ x_1 = \frac{3}{2} \end{cases}$$

The intersection point between those two lines is $D(\frac{3}{2}, 1)$.
To conclude with the intersections computations, I calculated the final intersection point between $x_2 \leq -2x_1 + 4$ and the x-axis , thus, the computations are the following:

$$\begin{cases} x_2 = -2x_1 + 4 \\ x_2 = 0 \end{cases} \rightarrow \begin{cases} x_2 = -2x_1 + 4 \\ 0 = -2x_1 + 4 \end{cases}$$
$$\begin{cases} x_2 = -2x_1 + 4 \\ 2x_1 = 4 \end{cases} \rightarrow \begin{cases} x_2 = -2 \cdot 2 + 4 \\ x_1 = 2 \end{cases}$$
$$\begin{cases} x_2 = 0 \\ x_1 = 2 \end{cases}$$

The intersection point between those two lines is $E(2, 0)$.
To sum up, the basic feasible points are $A(0, 0)$, $B(0, \frac{3}{2})$, $C(\frac{3}{13}, \frac{24}{13})$, $D(\frac{3}{2}, 1)$, $E(2, 0)$ and the origin $(0, 0)$; Now, we need to identify which point from the set is the one that minimized the function in order to resolve the optimization problem, thus, I evaluate the objective function at those vertices and pick the minimum value:

$$\begin{aligned}
&\text{at vertix } A(0,0) : f(A) = 4 \cdot 0 + 3 \cdot 0 = 0 \\
&\text{at vertix } B(0, \frac{3}{2}) : f(B) = 4 \cdot 0 + 3 \cdot \frac{3}{2} = \frac{9}{2} = 4.5 \\
&\text{at vertix } C(\frac{3}{13}, \frac{24}{13}) : f(C) = 4 \cdot \frac{3}{13} + 3 \cdot \frac{24}{13} = \frac{12}{13} + \frac{72}{13} = \frac{84}{13} \approx 6.46 \\
&\text{at vertix } D(\frac{3}{2}, 1) : f(D) = \cancel{4}^2 \cdot \frac{3}{\cancel{2}^1} + 3 \cdot 1 = 6 + 3 = 9 \\
&\text{at vertix } E(2,0) : f(E) = 4 \cdot 2 + 3 \cdot 0 = 8
\end{aligned}$$

In conclusion, it can be noticed that the point $A(0, 0)$ is the coordinate that minimize the objective function (5), then , $x_1 = 0$ and $x_2 = 0$ is the optimal solution for the linear programming problem. Next, I added the graph of the feasible region with the set of basic optimal solutions:

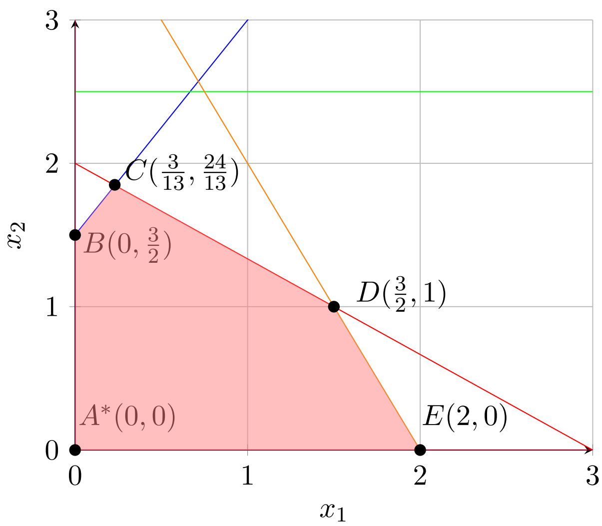

Figure 2: Feasible region for the problem (5)

c) Prove that the first order necessary conditions holds for the optimal point.
To begin with the verification of the first order necessary conditions, as first task, we need to remark the KKT conditions: the general case for these conditions has been explained in the ***Theorem 12.1*** in the previous exercise; in particular, KKT conditions for this constrained optimization can be described as following:

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\
c_i(x^*) &\geq 0, \quad \text{for all } i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, \quad \text{for all } i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.
\end{aligned}$$

where $\mathcal{E} = \emptyset$, $\mathcal{I} = \{0, 1, 2, 3, 4, 5\}$ in which $i = 0, 1, 2, 3, 4, 5$ correspond to the inequality constraints in the problem.
After this remark, it is necessary to compute the gradient of the objective function and the active constraints in order to evaluate the KKT conditions at the optimal point $A(0,0)$. Remark: the active set of constraints in which given the feasible point x^* the conditions are $c_i(x^*) = 0$,mathematically $\mathcal{A} = \mathcal{E} \cup \{i \in \mathcal{I} : c_i(x^*) = 0\}$,thus, the gradient of the objective function and the constraints are calculated as follows:

$$\nabla f(x) = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

Now, let us find the active constraints at the point $A(0,0)$, hence:

$$\begin{cases}
6 - 2 \cdot 0 - 3 \cdot 0 & \rightarrow 6 \geq 0 \\
3 + 3 \cdot 0 - 2 \cdot 0 & \rightarrow 3 \geq 0 \\
5 - 2 \cdot 0 & \rightarrow 5 \geq 0 \\
4 - 2 \cdot 0 - 0 & \rightarrow 4 \geq 0 \\
0 \geq 0 & \rightarrow 0 = 0 \\
0 \geq 0 & \rightarrow 0 = 0
\end{cases}$$

Then, the active constraints at the point $A(0,0)$ are the constraints c_4 and c_5,thus, $\mathcal{A} = \{4, 5\}$.
Subsequently, we can compute and substitute the optimal point in the active constraint gradients as follow:

$$\nabla c_4(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\nabla c_5(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Notice that, the LICQ property is satisfied at the point $A(0,0)$ because the gradients of the active constraints are not linearly dependent between each other at that specific point. Next, we need to check that $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$, thus it is possible to write as follows:

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= \nabla f(x) - \lambda_4 \nabla c_4(x) - \lambda_5 \nabla c_5(x) \\
&= \begin{bmatrix} 4 \\ 3 \end{bmatrix} - \lambda_4 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \lambda_5 \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{aligned}$$

It can easily be noticed that the gradient of the Lagrange function is equal to zero at the optimal point A when Lagrangian multipliers associated to the active constraints are $\lambda_4 = 4$ and $\lambda_5 = 3$. In conclusion, it is possible to affirm that the first order necessary conditions hold for the optimal point $A(0, 0)$.

References

1. Jorge Nocedal and Stephen J Wright. Numerical optimization. Springer, 1999
2. "Notes on the optimization" available on iCorsi platform