



Student: Jeferson Morales Mariciano

Discussed with:

Assignment 4

Due date: Monday, 3 June 2024, 12:00 AM

1. Exercise (20/100)

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

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

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

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

1. Answer

The function written in canonical form correspond to:

$$\begin{aligned} f(\mathbf{x}) &= 7x^2 + 4xy + y^2 \\ &= [7x + 2y \quad 2x + y] \begin{bmatrix} x \\ y \end{bmatrix} \\ &= [x \quad y] \begin{bmatrix} 7 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \frac{1}{2} [x \quad y] \begin{bmatrix} 14 & 4 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} \end{aligned}$$

With $\mathbf{b} = \mathbf{0}$, $c = 0$, and A being clearly a symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 14 & 4 \\ 4 & 2 \end{bmatrix}$$

Let's verify if A is positive definite as required by the quadratic form:

$$\begin{aligned}
\det(\lambda I - \mathbf{A}) &= \begin{vmatrix} \lambda - 14 & -4 \\ -4 & \lambda - 2 \end{vmatrix} \\
&= \lambda^2 - 16 + 12 \\
&\Rightarrow \lambda_{1,2} = 8 \pm 2\sqrt{13} > 0
\end{aligned}$$

Finally, since all eigenvalues are positive, \mathbf{A} is SPD.

2. Answer

The CG method is an iterative algorithm for solving a linear system of equations $Ax = b$ where $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. whose Hessian matrix is symmetric and **positive definite**. The performance of the linear CG method is determined by the **distribution of the eigenvalues** of the coefficient matrix, which are 2 so is a good candidate already. The CG method is appropriate and effective for minimizing the quadratic function.

2. Exercise (20/100)

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

$$\begin{aligned}
\min_{\mathbf{x}} f(\mathbf{x}) &:= -3x^2 + y^2 + 2z^2 + 2(x + y + z) \\
\text{subject to } c(\mathbf{x}) &= x^2 + y^2 + z^2 - 1 = 0
\end{aligned} \tag{2}$$

Write down the Lagrangian function and derive the KKT conditions for (2).

Answer

The constrained optimization problem can be written as:

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \quad \text{subject to} \quad \begin{cases} c_i(\mathbf{x}) = 0 & i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0 & i \in \mathcal{I} \end{cases} \tag{3}$$

where the **objective function** f and the **constraint functions** on the variables c_i are all smooth and real-valued defined on a subset of \mathbb{R}^n . The problem defines two finite sets of indices: \mathcal{I} for the **equality constraints** and \mathcal{E} for the **inequality constraints**. In addition, the set of points \mathbf{x} that satisfy the constraints is defined as the **feasible region** Ω :

$$\Omega = \{\mathbf{x} \mid c_i(\mathbf{x}) = 0, i \in \mathcal{E}; c_i(\mathbf{x}) \geq 0, i \in \mathcal{I}\} \tag{4}$$

Allowing to coincisely write the constrained optimization problem as:

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}) \tag{5}$$

Then, the **active set** $\mathcal{A}(\mathbf{x})$ at any feasible \mathbf{x} consists of the equality constraints indices from \mathcal{E} together with the indices of the inequality constraints i for which $c_i(\mathbf{x}) = 0$:

$$\mathcal{A}(\mathbf{x}) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(\mathbf{x}) = 0\} \tag{6}$$

So, at a feasible point \mathbf{x} , the inequality constraint $i \in \mathcal{I}$ is said to be active if $c_i(\mathbf{x}) = 0$ and inactive if the strict inequality $c_i(\mathbf{x}) > 0$ is satisfied.

Assuming a single equality scenario part of the active set, at the solution \mathbf{x}^* , the constraint normal $\nabla c_1(\mathbf{x}^*)$ is parallel to $\nabla f(\mathbf{x}^*)$, meaning that there is a scalar λ_1^* called **Lagrangian multiplier** such that:

$$\nabla f(\mathbf{x}^*) = \lambda_1^* \nabla c_1(\mathbf{x}^*) \tag{7}$$

Finally, the **Langragian function** and its gradient for the general problem are defined as:

$$\begin{aligned}
\mathcal{L}(\mathbf{x}, \lambda) &= f(\mathbf{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\mathbf{x}) \\
\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) &= \nabla f(\mathbf{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla c_i(\mathbf{x})
\end{aligned} \tag{8}$$

If assuming a single equality constraint scenario part of the active set, note that $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda_1) = \nabla f(\mathbf{x}) - \lambda_1 \nabla c_1(\mathbf{x})$, allowing to write the condition (7) equivalently as follows:

$$\text{at solution } \mathbf{x}^*, \exists \lambda_1^* : \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda_1^*) = \mathbf{0} \tag{9}$$

This observation suggests that we can search for solutions of the equality-constrained problem (2) by seeking stationary points of the Lagrangian function. The conditions (7) and (9) are equivalent and necessary conditions for an optimal solution of the problem (2), but clearly not sufficient.

An important constraint qualification condition is **LICQ**: given the point \mathbf{x} and the active set $\mathcal{A}(\mathbf{x})$ defined in (6), we say that the Linear Independence Constraint Qualification (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(\mathbf{x}) \mid i \in \mathcal{A}(\mathbf{x})\}$ is linearly independent. In general, if LICQ holds, none of the active constraints gradients can be zero.

The following necessary conditions defined are called first-order conditions because they are concerned with properties of the gradients (first-derivative vectors) of the objective and constraint functions. **First-Order Necessary Conditions**: suppose that \mathbf{x}^* is a local solution of (2), that the functions f and c_i in (2) are continuously differentiable, and that the LICQ holds at \mathbf{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 $(\mathbf{x}^*, \lambda^*)$:

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) &= \mathbf{0} \\
c_i(\mathbf{x}^*) &= \mathbf{0} \quad \forall i \in \mathcal{E} \\
c_i(\mathbf{x}^*) &\geq \mathbf{0} \quad \forall i \in \mathcal{I} \\
\lambda_i^* &\geq 0 \quad \forall i \in \mathcal{I} \\
\lambda_i^* c_i(\mathbf{x}^*) &= \mathbf{0} \quad \forall i \in \mathcal{E} \cup \mathcal{I} \quad (\text{complementary conditions})
\end{aligned} \tag{10}$$

The conditions (10) are often known as the Karush-Kuhn-Tucker conditions, or **KKT conditions** for short. The last row in (10) contains the complementary conditions, they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both. In particular, the Lagrange multipliers corresponding to inactive inequality constraints are zero, we can omit the terms for indices $i \notin \mathcal{A}(\mathbf{x}^*)$ from the first condition in (10) and rewrite it as:

$$\mathbf{0} = \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) - \sum_{i \in \mathcal{A}(\mathbf{x}^*)} \lambda_i^* \nabla c_i(\mathbf{x}^*) \tag{11}$$

The derived Lagrangian function for the problem (2) is:

$$\begin{aligned}
\mathcal{L}(\mathbf{x}, \lambda) &= f(\mathbf{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\mathbf{x}) \quad \mathcal{E} = \{1\}, \mathcal{I} = \emptyset \\
&= f(\mathbf{x}) - \lambda c(\mathbf{x}) \\
&= -3x^2 + y^2 + 2z^2 + 2(x + y + z) - \lambda(x^2 + y^2 + z^2 - 1) \\
&= (-3 - \lambda)x^2 + (1 - \lambda)y^2 + (2 - \lambda)z^2 + 2(x + y + z) + \lambda
\end{aligned}$$

Let $\mathbf{x}^* = [x^*, y^*, z^*]^T$ be a local solution, then the derived KKT conditions are:

$$\begin{aligned}\nabla f(\mathbf{x}^*) &= \begin{bmatrix} -6x^* + 2 \\ 2y^* + 2 \\ 4z^* + 2 \end{bmatrix}, \quad \nabla c = \begin{bmatrix} 2x^* \\ 2y^* \\ 2z^* \end{bmatrix} \\ \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = \mathbf{0} &\Rightarrow \begin{bmatrix} -6x^* + 2 - 2\lambda^*x^* \\ 2y^* + 2 - 2\lambda^*y^* \\ 4z^* + 2 - 2\lambda^*z^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ c(\mathbf{x}^*) = 0 &\Rightarrow (x^*)^2 + (y^*)^2 + (z^*)^2 - 1 = 0 \\ \lambda^* c(\mathbf{x}^*) = 0 &\Rightarrow \lambda^* ((x^*)^2 + (y^*)^2 + (z^*)^2 - 1) = 0\end{aligned}$$

Note that if equality condition $c(\mathbf{x}^*) = 0$ holds, then condition $\lambda^* c(\mathbf{x}^*) = 0$ is also satisfied.

3. Exercise (60/100)

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 constrained minimization problem, $\mathbf{x} = (x_1, x_2)^T$;

$$\min_{\mathbf{x}} f(\mathbf{x}) := 4x_1 + 3x_2 \quad (12)$$

subject to:

$$\begin{aligned}6 - 2x_1 - 3x_2 &\geq 0 \\ 3 + 3x_1 - 2x_2 &\geq 0 \\ 5 - 2x_2 &\geq 0 \\ 4 - 2x_1 - x_2 &\geq 0 \\ x_2 &\geq 0 \\ x_1 &\geq 0\end{aligned} \quad (13)$$

- a) Sketch the feasible region for this problem.
- b) Which are the basic feasible points of the problem 12? Compute them by hand using the geometrical interpretation and find the optimal point \mathbf{x}^* that minimizes f subject to the constraints.
- c) Prove that the first order necessary conditions holds for the optimal point.

1. Answer

The Simplex method is ...

The algorithm works by ...

Algorithm block ...

The search direction ...

2. Answer

a)

The sketch of the feasible region in Figure 1 was done with an online tool using geogebra as backend.

b)

c)

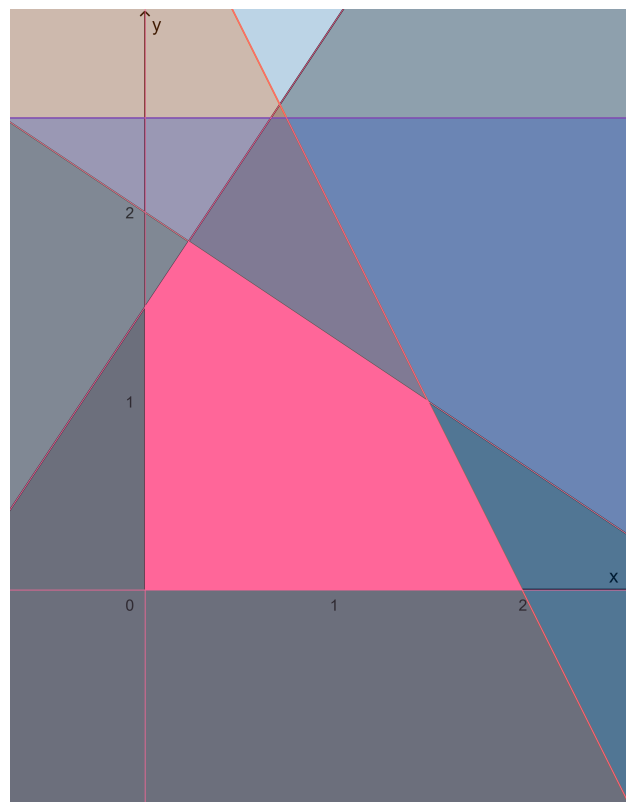


Figure 1: Feasible region for the problem 12 and 13