# TMA947 Assignment Exercises 5

Edoardo Mangia

September 2024

## Contents



## 1 HW 1.2

### Problem Statement

Determine whether the direction

$$\mathbf{p} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

is a descent direction with respect to the function $f : \mathbb{R}^2 \to \mathbb{R}$, given by

$$f(\mathbf{x}) = x_2 e^{x_1} + x_1^2 - \cos(x_2) + 6,$$

at the point

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Provide a detailed explanation to support the conclusion.

### Solution

To determine if the direction $\mathbf{p}$ is a descent direction for the function $f$ at the point $\mathbf{x}$, the directional derivative of $f$ in the direction $\mathbf{p}$ must be evaluated. A direction $\mathbf{p}$ is considered a descent direction if the directional derivative is negative.

**Step 1: Compute the Gradient of $f$**

The gradient of $f$ is given by:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right).$$

Compute $\frac{\partial f}{\partial x_1}$:

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}\left(x_2 e^{x_1} + x_1^2 - \cos(x_2) + 6\right) = x_2 e^{x_1} + 2x_1.$$

Compute $\frac{\partial f}{\partial x_2}$:

$$\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2}\left(x_2 e^{x_1} + x_1^2 - \cos(x_2) + 6\right) = e^{x_1} + \sin(x_2).$$

Thus, the gradient is:

$$\nabla f(\mathbf{x}) = \left(x_2 e^{x_1} + 2x_1,\ e^{x_1} + \sin(x_2)\right).$$

**Step 2: Evaluate the Gradient at $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$**

Substitute $x_1 = 1$ and $x_2 = 0$ into the gradient:

$$\nabla f(1,0) = \left(0 \cdot e^1 + 2 \cdot 1,\ e^1 + \sin(0)\right) = (2, e).$$

**Step 3: Compute the Directional Derivative**

The directional derivative of $f$ at $\mathbf{x}$ in the direction $\mathbf{p}$ is given by:

$$D_{\mathbf{p}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{p}.$$

Compute $D_{\mathbf{p}} f(1,0)$:

$$D_{\mathbf{p}} f(1,0) = (2, e) \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 2 \cdot (-1) + e \cdot 1 = -2 + e.$$

Given that $e \approx 2.718$:

$$-2 + e \approx -2 + 2.718 = 0.718.$$

**Conclusion**

Since $D_{\mathbf{p}} f(1,0) \approx 0.718 > 0$, the directional derivative in the direction $\mathbf{p}$ is positive. Therefore, $\mathbf{p}$ is **not** a descent direction at the point $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

## 2 E 1.2

### Problem Statement

Weierstrass' Theorem states that an optimal solution of a continuous objective function always exists if either the feasible set is closed and bounded (compact) or if the set is closed and the objective is coercive. Consider the following problems: Determine whether an optimal solution exists for each problem. If not, explain how the problem violates Weierstrass' Theorem. (Hint: visualize the problems by drawing them.)

### Solution

To assess whether an optimal solution exists for each problem, the conditions of Weierstrass' Theorem must be evaluated:

- **Closed Set:** The feasible set contains all its boundary points. If a sequence of feasible points converges to a limit point, that limit point must also be feasible.
- **Bounded Set:** The feasible set is contained within some finite region in space, meaning there is an upper limit to how far feasible points can be from the origin or a reference point.
- **Coercive Function:** A function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive if:

  $$\lim_{\|x\| \to \infty} f(x) = \infty.$$

  This implies that $f$ grows without bound as $x$ moves away from the origin.

#### Problem Analysis

1. **Problem (a):**
   - **Optimality:** No optimal solution exists.
   - **Reason:** The feasible set is not closed.
2. **Problem (b):**
   - **Optimality:** No optimal solution exists.
   - **Reason:** The feasible set is not bounded.
3. **Problem (c):**
   - **Optimality:** No optimal solution exists.
   - **Reason:** The feasible set is not closed.
4. **Problem (d):**
   - **Optimality:** No optimal solution exists.
   - **Reason:** The objective function is not continuous on the feasible set.

#### Conclusion

For each problem analyzed, an optimal solution does not exist because they violate the conditions set by Weierstrass' Theorem:

- Problems (a) and (c) have feasible sets that are not closed.
- Problem (b) has a feasible set that is not bounded.
- Problem (d) has an objective function that is not continuous on the feasible set.

# 3 E 2.5

## Problem Statement

Consider the optimization problem:

$$\text{minimize} \quad (x+1)^2 + (y-2)^2 + (z+1)^2$$

subject to

$$\begin{aligned} x^2 + y^2 &\leq 1, \\ x &\geq 0, \\ y &\geq 0, \\ z &\geq 0. \end{aligned}$$

Determine whether the point $(0,1,0)^T$ is an optimal solution to this problem. Provide a comprehensive explanation.

## Solution

The objective function represents the squared Euclidean distance from any point $(x, y, z)$ to the center $(-1, 2, -1)$ in three-dimensional space. The goal is to find the point within the feasible region that minimizes this distance.

#### Objective Function Analysis

$$f(x,y,z) = (x+1)^2 + (y-2)^2 + (z+1)^2$$

This function is convex as it is a sum of convex quadratic functions.

#### Feasible Region Description

The constraints define the feasible region:

- $x^2 + y^2 \leq 1$: A solid cylinder of radius 1 centered along the $z$-axis.
- $x \geq 0$, $y \geq 0$, $z \geq 0$: Restrict the feasible region to the first octant.

**Candidate Point Evaluation**

Evaluate the objective function at $(0, 1, 0)$:

$$f(0, 1, 0) = (0 + 1)^2 + (1 - 2)^2 + (0 + 1)^2 = 1 + 1 + 1 = 3.$$

**Optimality Conditions**

Using the properties of convex optimization and the method of Lagrange multipliers, the point $(0, 1, 0)$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions to be optimal.

**Gradient of $f$**

$$\nabla f(x, y, z) = (2(x + 1),\ 2(y - 2),\ 2(z + 1)).$$

At $(0, 1, 0)$:

$$\nabla f(0, 1, 0) = (2(1),\ 2(-1),\ 2(1)) = (2, -2, 2).$$

**Lagrangian Formulation**

$$\mathcal{L}(x, y, z, \lambda, \mu_x, \mu_y, \mu_z) = (x + 1)^2 + (y - 2)^2 + (z + 1)^2 + \lambda(1 - x^2 - y^2) - \mu_x x - \mu_y y - \mu_z z.$$

**KKT Conditions**

1. **Stationarity:**

$$\nabla f(x, y, z) = \lambda \nabla(-x^2 - y^2) + \mu$$

   Which leads to:

$$2(x + 1) = -2\lambda x - \mu_x,$$
$$2(y - 2) = -2\lambda y - \mu_y,$$
$$2(z + 1) = -\mu_z.$$

2. **Primal Feasibility:**

$$x^2 + y^2 \leq 1,$$
$$x \geq 0,\ y \geq 0,\ z \geq 0.$$

3. **Dual Feasibility:**

$$\lambda \geq 0,\ \mu_x \geq 0,\ \mu_y \geq 0,\ \mu_z \geq 0.$$

4. **Complementary Slackness:**

$$\lambda(1 - x^2 - y^2) = 0,$$
$$\mu_x x = 0,$$
$$\mu_y y = 0,$$
$$\mu_z z = 0.$$

**Applying KKT Conditions at** $(0, 1, 0)$

1. **Stationarity:**

$$2(0+1) = -2\lambda \cdot 0 - \mu_x \quad \Rightarrow \quad 2 = -\mu_x \quad \Rightarrow \quad \mu_x = -2.$$

$$2(1-2) = -2\lambda \cdot 1 - \mu_y \quad \Rightarrow \quad -2 = -2\lambda - \mu_y.$$

$$2(0+1) = -\mu_z \quad \Rightarrow \quad 2 = -\mu_z \quad \Rightarrow \quad \mu_z = -2.$$

2. **Dual Feasibility:**

$$\lambda \geq 0, \;\; \mu_x = -2 < 0, \;\; \mu_y \geq 0, \;\; \mu_z = -2 < 0.$$

   The dual variables $\mu_x$ and $\mu_z$ violate the dual feasibility conditions as they are negative.

3. **Conclusion from KKT Conditions:**

   Since $\mu_x$ and $\mu_z$ are negative, which violates the dual feasibility conditions, the point $(0, 1, 0)$ does not satisfy the KKT conditions necessary for optimality.

**Final Conclusion**

The point $(0, 1, 0)^T$ is **not** an optimal solution to the given optimization problem because it fails to satisfy the dual feasibility conditions of the KKT theorem.

# 4 E 2.8

## Problem Statement

Consider the optimization problem:

$$f(x) = x^T A x$$

where:

$$A = \begin{pmatrix} 0.5 & 2 \\ 0 & 0.5 \end{pmatrix}$$

and the initial point:

$$x_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Perform the following tasks:

1. Apply the Steepest Descent method with Exact Line Search.
2. Apply Newton's Method.
3. Discuss whether a global optimal solution exists.

## Solution

### Part (a): Steepest Descent with Exact Line Search

**Step 1: Compute the Gradient of $f(x)$** For a quadratic function $f(x) = x^T Ax$, the gradient is:

$$\nabla f(x) = (A + A^T)x.$$

Compute $A^T$:

$$A^T = \begin{pmatrix} 0.5 & 0 \\ 2 & 0.5 \end{pmatrix}.$$

Thus, the gradient is:

$$\nabla f(x) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} x.$$

At the initial point $x_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$:

$$\nabla f(x_0) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}.$$

**Step 2: Determine the Steepest Descent Direction** The steepest descent direction is the negative of the gradient:

$$\mathbf{d}_0 = -\nabla f(x_0) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

**Step 3: Perform Exact Line Search** The exact line search seeks to find the optimal step size $\alpha$ that minimizes $f(x_0 + \alpha \mathbf{d}_0)$.

Compute:

$$x(\alpha) = x_0 + \alpha \mathbf{d}_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \alpha \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2\alpha \\ -1 + \alpha \end{pmatrix}.$$

Substitute into $f(x)$:

$$f(x(\alpha)) = (2\alpha, -1 + \alpha) \begin{pmatrix} 0.5 & 2 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} 2\alpha \\ -1 + \alpha \end{pmatrix}.$$

Expanding this, we obtain:

$$f(x(\alpha)) = 0.5(2\alpha)^2 + 2(2\alpha)(-1 + \alpha) + 0(-1 + \alpha)2\alpha + 0.5(-1 + \alpha)^2.$$

Simplifying:

$$f(x(\alpha)) = 2\alpha^2 - 4\alpha + 4\alpha^2 + 0 + 0.5(\alpha^2 - 2\alpha + 1) = 6.5\alpha^2 - 5\alpha - 0.5.$$

To minimize, take the derivative with respect to $\alpha$ and set it to zero:

$$\frac{df}{d\alpha} = 13\alpha - 5 = 0 \quad \Rightarrow \quad \alpha = \frac{5}{13}.$$

**Step 4: Update the Point** The updated point is:

$$x_1 = x_0 + \alpha \mathbf{d}_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \frac{5}{13} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{10}{13} \\ -\frac{8}{13} \end{pmatrix}.$$

**Part (b): Newton's Method**

**Step 1: Compute the Gradient of $f(x)$** As previously computed:

$$\nabla f(x) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} x.$$

At $x_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$:

$$\nabla f(x_0) = \begin{pmatrix} -2 \\ -1 \end{pmatrix}.$$

**Step 2: Compute the Hessian of $f(x)$** For a quadratic function $f(x) = x^T A x$, the Hessian is:

$$H_f(x) = A + A^T = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

**Step 3: Newton's Update Step** The Newton update step is:

$$x_1 = x_0 - H_f(x)^{-1} \nabla f(x_0).$$

First, compute the inverse of the Hessian:

$$H_f(x) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \det(H_f) = 1 \cdot 1 - 2 \cdot 2 = 1 - 4 = -3.$$

$$H_f(x)^{-1} = \frac{1}{-3} \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{pmatrix}.$$

Next, compute the product:

$$H_f(x)^{-1} \nabla f(x_0) = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Thus, the new point is:

$$x_1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

### Part (c): Existence of a Global Optimal Solution

The analysis concludes that:

- The Steepest Descent method, after one iteration, moves to the point $\left(\frac{10}{13}, -\frac{8}{13}\right)$.
- Newton's Method, after one iteration, reaches the point $(0, 0)$, which is a saddle point.
- The function $f(x) = x^T Ax$ is convex if and only if the Hessian $A + A^T$ is positive semidefinite. In this case:
$$A + A^T = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},$$
which has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -1$. Since one eigenvalue is negative, the function $f(x)$ is not convex.
- Due to the presence of a saddle point and the non-convexity of $f(x)$, a global optimal solution may not exist, or multiple local minima may exist without a unique global minimum.

### Final Conclusion

1. **Steepest Descent Method:** The method proceeds to $\left(\frac{10}{13}, -\frac{8}{13}\right)$ after one iteration with a step size of $\alpha = \frac{5}{13}$.
2. **Newton's Method:** The method converges to the saddle point $(0, 0)$ in a single iteration, indicating potential issues with non-convexity.
3. **Global Optimal Solution:** Due to the non-convexity of $f(x)$, a global optimal solution is not guaranteed. The presence of a saddle point suggests that the optimization landscape is more complex, and multiple local minima may exist without a unique global minimum.

**Answer:** A global optimal solution does not necessarily exist because the function $f(x) = x^T Ax$ is not convex, as the Hessian $A + A^T$ has both positive and negative eigenvalues. Consequently, the optimization landscape contains saddle points, and methods like Newton's may converge to non-optimal points.