HOMEWORKS # 4 & 5
SOLUTIONS

Ahmad F. Taha

October 14, 2023

Solutions.

---

1. **Installing and playing with CVX**: The objective of this exercise is to get you started with CVX—the convex optimization solver on Matlab. Do the following:
   - Go through `http://cvxr.com/cvx/` and `http://web.cvxr.com/cvx/doc/intro.html#what-is-cvx`
   - Download and Install CVX on Matlab: `http://cvxr.com/cvx/download/`, `http://web.cvxr.com/cvx/doc/install.html`
   - Read the first four chapters pages of the CVX User's Guide: `http://web.cvxr.com/cvx/doc/CVX.pdf` (till Page 25 of the PDF, it's not much I promise).

Solutions.

Hope you did install CVX. And to quote from King's *The Shawshank Redemption*, *hope is a good thing, maybe the best of things.*

---

2. Is the set $\mathcal{S}$ of all matrices of the form $\begin{bmatrix} 2a & b \\ 3a+b & 3b \end{bmatrix}$ a subspace of $\mathbb{R}^{2\times 2}$?

Solutions.

Yes. It indeed is a subspace. Take two matrices $S_1 = \begin{bmatrix} 2a_1 & b_1 \\ 3a_1+b_1 & 3b \end{bmatrix}$ and $S_2 = \begin{bmatrix} 2a_2 & b_2 \\ 3a_2+b_2 & 3b_2 \end{bmatrix}$ where both these matrices are in $\mathcal{S}$.

Then you can see that

$$\lambda_1 S_1 + \lambda_2 S_2 = S_3 = \begin{bmatrix} a_3 & b_3 \\ 3a_3+b_3 & 3b_3 \end{bmatrix} \in \mathcal{S}$$

for any choice of $\lambda_{1,2}$, meaning that the resulting matrix maintains the structure of all matrices in $\mathcal{S}$.

---

3. Is $\mathcal{S} = \left\{ \begin{bmatrix} a+2b \\ a+1 \\ a \end{bmatrix}; a, b \in \mathbb{R} \right\}$ a subspace of $\mathbb{R}^3$?

Solutions.

It is *not* a subspace of $\mathbb{R}^3$. Similar to the previous problems, take two vectors S_1 and S_2 in $\mathcal{S}$. You can see that the addition of these two vectors results in a vector has an additional constant of 2 in the second entry of the vector. You can prove this via a counter example.

Let $S_1 = \begin{bmatrix} 2+2\cdot 1 \\ 2+1 \\ 2 \end{bmatrix}$ $(a_1 = 2, b_1 = 1)$ and $S_2 = \begin{bmatrix} 1+2\cdot 2 \\ 1+1 \\ 1 \end{bmatrix}$ $(a_2 = 1, b_2 = 2)$, then

$$S_3 = S_1 + S_2 = \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix}.$$

For vector $\begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = S_3$ to be in $\mathcal{S}$, and since $5 = 5 = a + 2b = a + 1$, then $b = 0.5$, meaning that $a + 1 = 5$ or $a = 4$. However, the last entry in S_3 is equal to 3, which is not equal to a, hence a contradiction and $S_3 \notin \mathcal{S}$.

4. You are given the following optimization problem

$$\begin{aligned} \text{minimize} \quad & 0.025P_{G_1}^2 + 20P_{G_1} + 0.05P_{G_2}^2 + 25P_{G_2} + \pi \\ \text{subject to} \quad & P_{G_1} + P_{G_2} = P_L \\ & P_{G_1} \geq 0,\ P_{G_2} \geq 0 \end{aligned}$$

(a) Solve the above optimization problem using CVX where $P_L = 250$. Report the values for the solutions.

(b) Plot the solutions for various $P_L = 150, 200, 250, 270, \ldots.$

(c) What happens as you change the coefficient for the quadratic terms in the objective function? That is, what does changing 0.025 and 0.05 in the objective function result in?

(d) Add an upper bound $P_{G_1} > 200$ to the optimization problem above and solve it again. Report the values.

(e) Does the constant π play any role in the solution to the above optimization problem? What is this role (if any)? And why? No math allowed. Just give me your intuition.

Solutions.

The problem can be generalized as a quadratic optimization problem, which can be written as in(1).

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^{\mathbf{T}} Q\mathbf{x} + q\mathbf{x} + r && (1) \\ \text{where} \quad & Q = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.10 \end{bmatrix} && (2) \\ & p = \begin{bmatrix} 20 & 25 \end{bmatrix} && (3) \\ & r = \pi && (4) \\ \text{subject to} \quad & \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x} = P_L && (5) \\ & \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x} \geq 0 && (6) \\ & \begin{bmatrix} 0 & 1 \end{bmatrix} \mathbf{x} \geq 0 && (7) \end{aligned}$$

Write the quadratic optimization in a CVX format as follows.

```
clear;clc;
Q = [0.05 0; 0 0.10];
q = [20 25];
r = pi;
%Dimension of the decision variable
n = 2;
%This term is flexible
L = 250;
cvx_begin
variable x(n);
minimize (0.5*quad_form(x,Q)+q*x+r);
subject to
[1 1]*x == L
[1 0]*x >= 0
[0 1]*x >= 0
cvx_end
```

(a) CVX result summary is posted as follows. $P_{G_1} = 200$ and $P_{G_2} = 50$ is the optimal solution, and the minimal objective function value is 6378.14.

```
Calling SDPT3 4.0: 6 variables
num. of constraints = 2
dim. of linear var = 2
Status: Solved
Status: Solved
Optimal value (cvx_optval): +6378.14
x = 2x1
200.0044
49.9956
```

(b) Figure(1) shows the change of P_L and the related solutions.

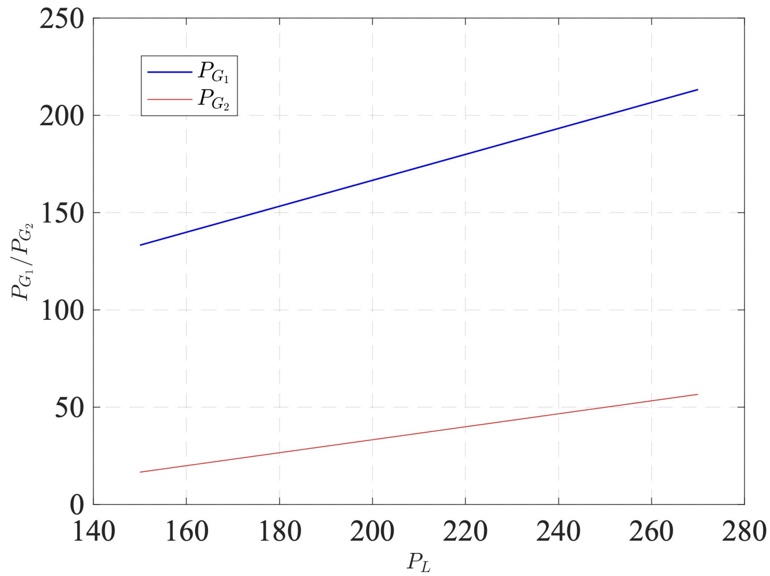


Figure 1: Solutions for various P_L setting

(c) The change of the coefficient for the quadratic terms will have an impact on the result of the decision variables as the matrix Q in Equation(1) will be different.

(d) The MATLAB code will be modified as follows if the upper bound $P_{G_1} > 200$ is added.

```
clear;clc;
Q = [0.05 0; 0 0.10];
q = [20 25];
r = pi;
%Dimension of the decision variable
n = 2;
%This term is flexible
L = 250;
cvx_begin
variable x(n);
minimize (0.5*quad_form(x,Q)+q*x+r);
subject to
[1 1]*x == L
% Modification term
[1 0]*x > 200
[0 1]*x >= 0
cvx_end
```

The result is nearly the same as the original one. With the increase of P_L, this constraint will make more difference in the result.

```
x = 2x1
200.0083
49.9917
```

(e) The constant π has no relationship with the solution to the above optimization problem. Because no matter what it is, the constant term is independent on the decision variables.

Figure 2: This is the CVX code and the corresponding screenshot.

5. Prove that the hypeplane $\mathcal{H} = \{x \in \mathbb{R}^n \mid a^T x = 0\}$ is a subspace.

Solutions.

For any $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ require: $\mathbf{a}^T\mathbf{x} = 0$ and $\mathbf{a}^T\mathbf{y} = 0$, for any $\lambda, \mu \in \mathbb{R}$, the following equations are satisfied.

$$\begin{cases} \lambda\mathbf{a}^T\mathbf{x} = 0 \\ \mu\mathbf{a}^T\mathbf{y} = 0 \end{cases} \tag{8}$$

Hence, $\mathbf{a}^T(\lambda\mathbf{x}) + \mathbf{a}^T(\mu\mathbf{y}) = 0$ is always satisfied. To extract common factors $\mathbf{a}$, the equation can be transformed as follows.

$$\mathbf{a}^T(\lambda\mathbf{x} + \mu\mathbf{y}) = 0 \tag{9}$$

Equation(9) indicates that the vector $\lambda\mathbf{x} + \mu\mathbf{y} \in \mathcal{H}$. The hyperplane $\mathcal{H} = \{x \in \mathbb{R}^n \mid a^T x = 0\}$ is hence a subspace, by satisfying the properties of subspaces.

---

6. Prove that the halfspace $\mathcal{S} = \{x \in \mathbb{R}^n \mid a^T x \leq 0\}$ is a convex cone.

Solutions.

For any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ require: $\mathbf{a}^T\mathbf{x} \leq 0$ and $\mathbf{a}^T\mathbf{y} \leq 0$, for any $\lambda, \mu \in \mathbb{R}$, the following equations are satisfied.

$$\begin{cases} \lambda \mathbf{a}^T\mathbf{x} \leq 0 \\ \mu \mathbf{a}^T\mathbf{y} \leq 0 \end{cases} \tag{10}$$

Hence, $\mathbf{a}^T(\lambda\mathbf{x}) \leq 0$ and $\mathbf{a}^T(\lambda\mathbf{x}) + \mathbf{a}^T(\mu\mathbf{y}) \leq 0$ is always satisfied, the halfspace $\mathcal{S}$ is a cone can be proved. To extract common factors $\mathbf{a}$, the equation can be transformed as follows.

$$\mathbf{a}^T(\lambda\mathbf{x} + \mu\mathbf{y}) \leq 0 \tag{11}$$

Equation(11) indicates that the vector $\lambda\mathbf{x} + \mu\mathbf{y} \in \mathcal{S}$. The hyperplane $\mathcal{H} = \{x \in \mathbb{R}^n \mid a^T x \leq 0\}$ is hence a convex cone by satisfying the properties of convex cones.

---

7. Show that a *slab, i.e.*, a set of the form $S = \{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\}$ is a convex set.

Solutions.

Many ways to prove this to be true, but here's one way to do so. For any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ require: $\alpha \leq \mathbf{a}^T\mathbf{x} \leq \beta$ and $\alpha \leq \mathbf{a}^T\mathbf{y} \leq \beta$, for any $\theta \in \mathbb{R}$, the following equations are satisfied.

$$\begin{cases} \theta\alpha \leq \theta\mathbf{a}^T\mathbf{x} \leq \theta\beta \\ (1-\theta)\alpha \leq (1-\theta)\mathbf{a}^T\mathbf{y} \leq (1-\theta)\beta \end{cases} \tag{12}$$

So we now need to prove that the point $z = \theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathcal{S}$ and that it saifies the condition of a slab. To that end, we can write:

$$\theta\alpha + (1-\theta)\alpha \leq \theta\mathbf{a}^T\mathbf{x} + (1-\theta)\mathbf{a}^T\mathbf{y} \leq \theta\beta + (1-\theta)\beta \tag{13}$$

by adding the corresponding LHS and RHS (and what's in between) of (12). This inequality can be reduced to:

$$\alpha \leq \mathbf{a}^T(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \beta \tag{14}$$

Equation(14) indicates that $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathcal{S}$, hence a slab is a convex set.

---

8. Show that the set Ω given by $\Omega = \{y \in \mathbb{R}^2; \|y\|_2^2 \leq 4\}$ is convex, where $\|y\|_2^2 = y^\top y$.

Hint: Show that if $z = \beta x + (1-\beta)y$, then $\|z\|_2^2 \leq 4$. You might find the submultiplicative matrix-vector property to be useful too.

Solutions.

The set Ω is convex if $x, y \in \Omega$, then $z = \beta x + (1-\beta)y$ is a point on the line joining x and y should also be in Ω, where $0 \leq \beta \leq 1$. Hence, the problem reduces to showing that $z = \beta x + (1-\beta)y \in \Omega$. Plugging in what z actually is, we obtain:

$$\begin{aligned}
\|z\|^2 &= z^\top z \\
\|z\|^2 &= (\beta x + (1-\beta)y)^\top (\beta x + (1-\beta)y) \\
&= \beta^2 \|x\|^2 + 2\beta(1-\beta)x^\top y + (1-\beta)^2\|y\|^2 \\
&\quad \text{via the submultiplicative matrix-vector property} \Rightarrow \\
&\leq \beta^2\|x\|^2 + 2\beta(1-\beta)\|x\|\|y\| + (1-\beta)^2\|y\|^2 \\
&\quad \text{via knowing that the norms of } x \text{ and } y \text{ can at most be } 2 \Rightarrow \\
&\leq 4\beta^2 + 8\beta(1-\beta) + 4(1-\beta)^2 \\
&= 4\beta^2 + 8\beta - 8\beta^2 + 4 + 4\beta^2 - 8\beta \\
&= 4
\end{aligned}$$

Hence $\|z\|^2 \leq 4$, which proves that z is in Ω and that Ω is a convex set.

---

9. For a positive semidefinite matrix $A = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_4 & x_5 \\ x_3 & x_5 & x_6 \end{bmatrix}$, give an explicit description of the positive semidefinite cone in terms of the matrix coefficients and ordinary inequalities.

Solutions.

The positive semi-definite cone means that the **principle minors** are nonnegative, then the equation can be derived as follows:

$$\begin{cases} x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \\ \det(\begin{bmatrix} x_1 & x_2 \\ x_2 & x_4 \end{bmatrix}) = x_1x_4 - x_2^2 \geq 0 \\ \det(\begin{bmatrix} x_1 & x_3 \\ x_3 & x_6 \end{bmatrix}) = x_1x_6 - x_3^2 \geq 0 \\ \det(\begin{bmatrix} x_4 & x_5 \\ x_5 & x_6 \end{bmatrix}) = x_4x_6 - x_5^2 \geq 0 \\ \det(\begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_4 & x_5 \\ x_3 & x_5 & x_6 \end{bmatrix}) = x_1x_4x_6 + 2x_2x_3x_5 - x_1x_5^2 - x_2^2x_6 - x_3^2x_4 \geq 0 \end{cases} \tag{15}$$

---

10. For the following function, find the set of values for β such that the function is convex.

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + 5x_3^2 - 2x_1x_3 + 2\beta x_1x_2 + 4x_2x_3.$$

Solutions.

The function $f(x_1, x_2, x_3)$ can be written as follows.

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + q\mathbf{x}^T + r \tag{16}$$

Here, $\mathbf{Q}$ can be written as follows.

$$\mathbf{Q} = \begin{bmatrix} 2 & 2\beta & -2 \\ 2\beta & 2 & 4 \\ -2 & 4 & 10 \end{bmatrix} \succeq \mathbf{0} \tag{17}$$

The function is convex when $\mathbf{Q}$ is semi-definite or definite, as stated in Equation(17). The following inequalities should be satisfied.

$$\begin{cases} 2 \geq 0 \\ 2 \geq 0 \\ 10 \geq 0 \\ \det(\begin{bmatrix} 2 & 2\beta \\ 2\beta & 2 \end{bmatrix}) = 4 - 4\beta^2 \geq 0 \\ \det(\begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix}) = 4 \geq 0 \\ \det(\begin{bmatrix} 2 & -2 \\ -2 & 10 \end{bmatrix}) = 16 \geq 0 \\ \det(\begin{bmatrix} 2 & 2\beta & -2 \\ 2\beta & 2 & 4 \\ -2 & 4 & 10 \end{bmatrix}) = 40 - 16\beta - 16\beta - 8 - 32 - 40\beta^2 \geq 0 \end{cases} \tag{18}$$

Arrange the inequalities well, then the following two inequalities can be calculated:

$$\begin{cases} \beta^2 \leq 1 \\ 5\beta^2 + 4\beta \leq 0 \end{cases} \tag{19}$$

As a result, the set of values for $\beta \in [-0.8, 0]$ entails that the function is convex in that range.

---

11. For each of the following functions determine whether it is convex, concave or neither:

    (a) $f(x) = e^x - 1$ on $\mathbb{R}$.

    (b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}_{++}$.

    Solutions.

    Second-order condition is applied in this case as the two functions are continuous in the defined domain. Meanwhile, the domain are convex sets.

    (a) $\nabla^2 f(x) = e^x > 0$ is always convex in $\mathbb{R}$. Hence, $f(x)$ is a strictly convex function as given function includes an affine transformation of e^x. But you can also use the second order derivative (e^x) which is also always positive, meaning the function itself is convex.

    (b) $\nabla^2 f(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is positive definite for all $x_1, x_2 \in \mathbb{R}^2_{++}$. Hence, $f(x_1, x_2)$ is a strictly convex function.

12. Consider the optimization problem:

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & 2x_1 + x_2 \geq 1 \\ & x_1 + 3x_2 \geq 1 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{array}$$

Make a sketch of the feasible set. For each of the following objective functions, give the optimal set and the optimal value using CVX.

(a) $f_0(x_1, x_2) = x_1 + x_2$.

(b) $f_0(x_1, x_2) = -x_1 - x_2$.

(c) $f_0(x_1, x_2) = x_1$.

(d) $f_0(x_1, x_2) = \max\{x_1, x_2\}$.

(e) $f_0(x_1, x_2) = x_1^2 + 9x_2^2$

Solutions.

(a) $x_1 = 0.4, \quad x_2 = 0.2, \quad f(x_1, x_2) = 0.6$

```
n=2
cvx_begin
variable x(n);
minimize [1 1]*x;
subject to
[2 1]*x >= 1
[1 3]*x >= 1
[1 0]*x >= 0
[0 1]*x >= 0
cvx_end
x
```

(b) $x_1 = 1.0, \quad x_2 = 0.0, \quad f(x_1, x_2) = -1$

```
n=2
cvx_begin
variable x(n);
minimize [-1 -1]*x;
subject to
[2 1]*x >= 1
[1 3]*x >= 1
[1 0]*x >= 0
[0 1]*x >= 0
cvx_end
x
```

(c) $x_1 = 0, \ \ x_2 = 6.0, \ \ f(x_1, x_2) = 0$

```
n=2
cvx_begin
variable x(n);
minimize [1 0]*x;
subject to
[2 1]*x >= 1
[1 3]*x >= 1
[1 0]*x >= 0
[0 1]*x >= 0
cvx_end
x
```

(d) $x_1 = 0.33, \ \ x_2 = 0.33, \ \ f(x_1, x_2) = 0.33$

```
n=2
cvx_begin
variable x(n);
minimize max(x);
subject to
[2 1]*x >= 1
[1 3]*x >= 1
[1 0]*x >= 0
[0 1]*x >= 0
cvx_end
x
```

(e) $x_1 = 0.500, \ \ x_2 = 0.167, \ \ f(x_1, x_2) = 0.5$

```
n=2
Q = [2 0; 0 18];
q = [0 0];
r = 0;
cvx_begin
variable x(n);
minimize (0.5*quad_form(x,Q)+q*x+r);
subject to
[2 1]*x >= 1
[1 3]*x >= 1
[1 0]*x >= 0
[0 1]*x >= 0
cvx_end
x
```

13. A first-order necessary optimality condition states the following:

*All points $x* \in \mathcal{C}$ satisfying $\nabla f_0(x*)^\top (x - x*) \geq 0$ for all $x \in \mathcal{C}$ will be **minimizers** of $f_0(x)$ over $\mathcal{C}$, when f_0 **is a convex function and** $\mathcal{C}$ **a convex set**.*

Use this condition to prove that $x^* = [1 \;\; 1/2 \;\; -1]^\top$ is optimal for the optimization problem. Once you're done, prove that this point is the optimal solution via a different approach (perhaps via solving the problem via the KKT conditions or any other method of your choice). Verify the solution via CVX.

$$\begin{array}{ll} \text{minimize} & 1/2 x^\top P x + q^\top x + r \\ \text{subject to} & -1 \leq x_i \leq 1, \;\; i = 1,2,3, \end{array}$$

where

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \quad r = 1.$$

Solutions.

Proof #1: According to the definition, calculate the 1st order derivative for the objective function $f_0(x)$:

$$\nabla f_0(\mathbf{x}) = \mathbf{P}\mathbf{x} + \mathbf{q} \tag{22}$$

As a result, $\nabla f_0(x^*)$ can be written as follows.

$$\nabla f_0(x^*) = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1}{2} \\ -1 \end{bmatrix} + \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix} \tag{23a}$$

$$= \begin{bmatrix} -1 & 0 & 2 \end{bmatrix}^\top \tag{23b}$$

Assume that $\mathbf{x} = [x_1 \; x_2 \; x_3]^\top$, then calculate the term $\nabla f_0(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*)$:

$$\nabla f_0(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) = -(x_1 - 1) + 2(x_3 + 1) \tag{24}$$

Considering that $-1 \leq x_i \leq 1$, the terms, $-(x_1 - 1) \geq 0$ and $2(x_3 + 1) \geq 0$. As a result, $\nabla f_0(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $x \in \mathcal{C}$. Accordingly, $f_0(x)$ is convex as the matrix $\mathbf{P}$ is definite positive. $\mathcal{C}$ is a convex set as the inequalities are all linear. Thus, according to the definition: x^* is the ***minimizer*** of $f_0(x)$ over $\mathcal{C}$.

Proof #2: KKT condition The KKT condition requires a Lagrangian function $\mathcal{L}(\mathbf{x}, \mu)$ for constraint optimization problems, $\mathbf{g}(\mathbf{x})$ refers to the inequality constraints of the optimization problem, $\mathbf{g}(\mathbf{x}) : \mathbb{R}^3 \to \mathbb{R}^6$.

$$\mathcal{L}(\mathbf{x}, \mu) = f_0(\mathbf{x}) + \mu^\top \mathbf{g}(\mathbf{x}) \tag{25}$$

$$\mu = [\mu_1 \; \mu_2 \; \mu_3 \; \mu_4 \; \mu_5 \; \mu_6]^\top \tag{26}$$

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \tag{27}$$

As a result, $\nabla \mathcal{L}(\mathbf{x}, \mu)$ can be calculated as follows and two equations for the KKT condition are stated.

$$\begin{cases} \nabla \mathcal{L}(\mathbf{x}, \mu) = \mathbf{P}\mathbf{x} + \mathbf{q} + \begin{bmatrix} -\mu_1 + \mu_2 & -\mu_3 + \mu_4 & -\mu_5 + \mu_6 \end{bmatrix}^\top = \mathbf{0} \\ \mu_i g_i(\mathbf{x}) = 0 \end{cases} \tag{28}$$

$$\mathbf{x} = -\begin{bmatrix} -1.68\mu_1 + 1.68\mu_2 + 1.56\mu_3 - 1.56\mu_4 - 1.06\mu_5 + 1.06\mu_6 - 0.56 \\ 1.56\mu_1 - 1.56\mu_2 - 1.52\mu_3 + 1.52\mu_4 + 1.02\mu_5 - 1.02\mu_6 - 0.98 \\ -1.06\mu_1 + 1.06\mu_2 + 1.02\mu_3 - 1.02\mu_4 - 0.77\mu_5 + 0.77\mu_6 + 1.48 \end{bmatrix} \tag{29}$$

- **Initial guess.** It is impossible that $\mu = \mathbf{0}$, as if $\mu = \mathbf{0}$, $\mathbf{x}^* = \begin{bmatrix} 0.56 & 0.98 & -1.48 \end{bmatrix}^\top \notin \mathcal{C}$.

- **Guess from the outlier.** Guess from the outlier one x_3, guess $x_3 = -1$, so $\mu_6 = 0$. Assume that $\mu_{1,2,3,4} = 0$, then $\mathbf{x}^* = \begin{bmatrix} 1.22 & 0.345 & -1 \end{bmatrix}^\top \notin \mathcal{C}$.

- **Guess again!** Guess from the outlier one x_1, guess $x_1 = 1$, so $\mu_1 = 0$. Assume that $\mu_{3,4} = 0$, then $\mu_2 = 1,\ \ \mu_5 = 2$, $\mathbf{x}^* = \begin{bmatrix} 1 & 0.5 & -1 \end{bmatrix}^\top \in \mathcal{C}$.

- **Check SONC.**
$$\nabla^2 \mathcal{L}(\mathbf{x}, \mu) = \mathbf{P} \succeq \mathbf{0} \tag{30}$$

Hence, $x^* = \begin{bmatrix} 1 & 0.5 & -1 \end{bmatrix}^\top$ is the optimal solution.

Verification: CVX

```
clear;clc;
P = [13 12 -2; 12 17 6; -2 6 12];
q = [-22 -14.5 13];
r = 1;
n = 3;
cvx_begin
variable x(n);
minimize (0.5*quad_form(x,P)+q*x+r);
subject to
[1 0 0; -1 0 0; 0 1 0; 0 -1 0; 0 0 1; 0 0 -1]*x ...
+[-1 ;-1; -1; -1; -1; -1] <= 0
cvx_end
x
```

CVX output is consistent with that of KKT one and the proved one.

```
Status: Solved
Optimal value (cvx_optval): -21.625

x =

1.0000
0.5000
-1.0000
```

14. Using the KKT conditions discussed in class, obtain all the candidate strict local minima for the following nonlinear optimization problem:

$$\begin{array}{ll} \max & -x_1^2 - 2x_2^2 \\ \text{subject to} & x_1 + x_2 \geq 3 \\ & x_2 - x_1^2 \geq 1 \end{array}$$

There are many cases to consider. Make sure that you don't miss any.

Solutions.

Transform the original optimization problem to the standard statement.

$$\begin{array}{ll} \min & x_1^2 + 2x_2^2 \\ \text{subject to} & -x_1 - x_2 + 3 \leq 0 \\ & x_1^2 - x_2 + 1 \leq 0 \end{array}$$

Write down the Lagrangian function of this optimization problem. That is:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mu) &= x_1^2 + 2x_2^2 + \mu_1(-x_1 - x_2 + 3) + \mu_2(x_1^2 - x_2 + 1) && (31) \\ &= (1 + \mu_2)x_1^2 - \mu_1 x_1 + 2x_2^2 - (\mu_1 + \mu_2)x_2 + 3\mu_1 + \mu_2 && (32) \end{aligned}$$

$$\begin{cases} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu) = \begin{bmatrix} 2(1+\mu_2)x_1 - \mu_1 \\ 4x_2 - (\mu_1 + \mu_2) \end{bmatrix} = \mathbf{0} \\ \mu_i g_i(\mathbf{x}) = 0 \end{cases} \tag{33}$$

$$\mathbf{x} = \begin{bmatrix} \frac{\mu_1}{2+2\mu_2} & \frac{\mu 1 + \mu_2}{4} \end{bmatrix}^\top \tag{34}$$

- $\mu_1 = \mu_2 = 0$, $\mathbf{x} = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$. In this case, inequality cannot be satisfied.
- $\mu_1 = 0$, $x_1^2 - x_2 + 1 = 0$. $\mathbf{x} = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$. In this case, inequality cannot be satisfied.
- $\mu_2 = 0$, $-x_1 - x_2 + 3 = 0$. $\mathbf{x} = \begin{bmatrix} -2 & -1 \end{bmatrix}^\top$. In this case, inequality cannot be satisfied.
- $x_1^2 - x_2 + 1 = 0$, $-x_1 - x_2 + 3 = 0$. The function has two solutions.
  (a) $\mathbf{x} = \begin{bmatrix} -2 & 5 \end{bmatrix}^\top$. $\mu = \begin{bmatrix} 28 & -8 \end{bmatrix}^\top$. In this case, inequality cannot be satisfied.
  (b) $\mathbf{x} = \begin{bmatrix} 1 & 2 \end{bmatrix}^\top$. $\mu = \begin{bmatrix} 6 & 2 \end{bmatrix}^\top$. In this case, inequality can be satisfied and $x^* = \begin{bmatrix} 1 & 2 \end{bmatrix}^\top$.
- **Check SONC.**

$$\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}, \mu) = \begin{bmatrix} 2(1+\mu_2) & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 4 \end{bmatrix} \succeq \mathbf{0} \tag{35}$$

Hence, $x^* = \begin{bmatrix} 1 & 2 \end{bmatrix}^\top$ is the candidate strict local minima for the problem.

---

15. Using the KKT conditions discussed in class, obtain all the candidate strict local minima for the following nonlinear optimization problem:

$$\begin{array}{ll} \min & x_1 + x_2^2 \\ \text{subject to} & x_1 - x_2 = 5 \\ & x_1^2 + 9x_2^2 \leq 36 \end{array}$$

There are many cases to consider. Make sure that you don't miss any.

Solutions.

Write down the Lagrangian function for this optimization problem. That is:

$$\mathcal{L}(\mathbf{x}, \mu, \lambda) = \mu x_1^2 + (\lambda + 1)x_1 + (9\mu + 1)x_2^2 - \lambda x_2 - 5\lambda - 36\mu \tag{36}$$

$$\begin{cases} \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mu) = \begin{bmatrix} 2\mu x_1 + (\lambda + 1) \\ 2(9\mu + 1)x_2 - \lambda \end{bmatrix} = \mathbf{0} \\ \mu(x_1^2 + 9x_2^2 - 36) = 0 \\ x_1 - x_2 - 5 = 0 \end{cases} \tag{37}$$

- $\mu = 0$, and $x_1 = 4.5,\ x_2 = -0.5, \lambda = -1$.
- $x_1^2 + 9x_2^2 - 36 = 0$, and $\mathbf{x}$ can be two solutions:
  (a) $\mathbf{x} = \begin{bmatrix} 5.572 & 0.572 \end{bmatrix}^\top$, $\mu = -0.183 < 0$. Then this solution will be dropped.
  (b) $\mathbf{x} = \begin{bmatrix} 3.428 & -1.572 \end{bmatrix}^\top$, $\mu = -0.100 < 0$ Then this solution will be dropped.
- **Check the SONC.**

$$\nabla^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mu, \lambda) = \begin{bmatrix} 2\mu & 0 \\ 0 & 2(9\mu + 1) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \succeq \mathbf{0} \tag{38}$$

Hence, $x^* = \begin{bmatrix} 4.5 & -0.5 \end{bmatrix}^\top$ is the candidate strict local minima for the problem

---

16. Consider the function

$$f(x_1, x_2) = 0.5x_1^2 + x_2^2 - x_1 - x_2 + 7$$

Use the Newton's method to minimize $f(x_1, x_2)$ with fixed time step size $t = 1$ and with initial guess $x^{(0)} = [0 \ \frac{1}{2}]^\top$.

Solutions.

$$f'(\mathbf{x}) = \begin{bmatrix} x_1 - 1 \\ 2x_2 - 1 \end{bmatrix} \tag{39}$$

$$f''(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \tag{40}$$

$$(\nabla^2 f(\mathbf{x}))^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \tag{41}$$

- For $\mathbf{x}^{(0)} = \begin{bmatrix} 0 & \frac{1}{2} \end{bmatrix}^\top$, $f(\mathbf{x}^{(0)}) = 6.75$, $f'(\mathbf{x}^{(0)}) = \begin{bmatrix} -1 & 0 \end{bmatrix}^\top$, then:

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - t(\nabla^2 f(\mathbf{x}^0))^{-1} f'(\mathbf{x}^{(0)}) = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}^\top \tag{42}$$

- For $\mathbf{x}^{(1)} = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}^\top$, $f(\mathbf{x}^{(0)}) = 6.25$, $f'(\mathbf{x}^{(1)}) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$, then:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - t(\nabla^2 f(\mathbf{x}^1))^{-1} f'(\mathbf{x}^{(1)}) = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}^\top \tag{43}$$

- $||\mathbf{x}^{(2)} - \mathbf{x}^{(1)}|| = 0$, the convergence criteria is satisfied.

Hence, $\mathbf{x}^{(*)} = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}^\top$ is the optimal solution to $f(\mathbf{x})$.

---

17. Consider an unknown function whose gradient is given as follows

$$\nabla f(x_1, x_2) = \begin{bmatrix} 8x_1 - 4x_2 \\ -4x_1 + 4x_2 \end{bmatrix}$$

Compute the first two iterations of the steepest descent algorithm for the above function with a given gradient. You can consider that the starting point is $x^{(0)} = [2 \quad 3]^\top$. Since $f(.)$ is unknown you are not allowed to use any closed form expression for $f(.)$. Use only its gradients. Test this algorithm with various step sizes t.

Once you are done, test different values for t (i.e., $t = 1, 0.95, 0.9, \ldots, 0.1$) via automating this procedure via a for loop. What is the optimal value of t that results in the best convergence towards a stationary point or a local minimum? A figure showcasing this would be nice.

Solutions.

(a) We follow the steepest descent algorithm for different values for t.

- $\mathbf{x}^{(0)} = \begin{bmatrix}2 & 3\end{bmatrix}^\top$, $\nabla f(\mathbf{x}^{(0)}) = \begin{bmatrix}4 & 4\end{bmatrix}^\top$.

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - t\nabla f(\mathbf{x}^{(0)}) = \begin{bmatrix}2-4t & 3-4t\end{bmatrix}^\top \tag{44}$$

  Assume $\nabla f(\mathbf{x}^{(1)}) = \mathbf{0}$, then $t = 0.25$.

- $\mathbf{x}^{(1)} = \begin{bmatrix}1 & 2\end{bmatrix}^\top$, $\nabla f(\mathbf{x}^{(1)}) = \begin{bmatrix}0 & 4\end{bmatrix}^\top$.

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - t\nabla f(\mathbf{x}^{(1)}) = \begin{bmatrix}1 & 2-4t\end{bmatrix}^\top \tag{45}$$

  Assume $\nabla f(\mathbf{x}^{(2)}) = \mathbf{0}$, then $t = 0.25$.

- $\mathbf{x}^{(2)} = \begin{bmatrix}1 & 1\end{bmatrix}^\top$, $\nabla f(\mathbf{x}^{(2)}) = \begin{bmatrix}4 & 0\end{bmatrix}^\top$.

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} - t\nabla f(\mathbf{x}^{(2)}) = \begin{bmatrix}1-4t & 1\end{bmatrix}^\top \tag{46}$$

  Assume $\nabla f(\mathbf{x}^{(3)}) = \mathbf{0}$, then $t = 0.125$.

- $\mathbf{x}^{(3)} = \begin{bmatrix}0.5 & 1\end{bmatrix}^\top$, $\nabla f(\mathbf{x}^{(3)}) = \begin{bmatrix}0 & 2\end{bmatrix}^\top$.

$$\mathbf{x}^{(4)} = \mathbf{x}^{(3)} - t\nabla f(\mathbf{x}^{(3)}) = \begin{bmatrix}0.5 & 1-2t\end{bmatrix}^\top \tag{47}$$

  Assume $\nabla f(\mathbf{x}^{(4)}) = \mathbf{0}$, then $t = 0.25$.

- $\mathbf{x}^{(4)} = \begin{bmatrix}0.5 & 0.5\end{bmatrix}^\top$, $\nabla f(\mathbf{x}^{(4)}) = \begin{bmatrix}2 & 0\end{bmatrix}^\top$.

$$\mathbf{x}^{(5)} = \mathbf{x}^{(4)} - t\nabla f(\mathbf{x}^{(4)}) = \begin{bmatrix}0.5-2t & 0.5\end{bmatrix}^\top \tag{48}$$

  Assume $\nabla f(\mathbf{x}^{(5)}) = \mathbf{0}$, then $t = 0.125$.

- $\mathbf{x}^{(5)} = \begin{bmatrix}0.25 & 0.5\end{bmatrix}^\top$, $\nabla f(\mathbf{x}^{(5)}) = \begin{bmatrix}0 & 1\end{bmatrix}^\top$.

$$\mathbf{x}^{(6)} = \mathbf{x}^{(5)} - t\nabla f(\mathbf{x}^{(5)}) = \begin{bmatrix}0.25 & 0.5-t\end{bmatrix}^\top \tag{49}$$

  Assume $\nabla f(\mathbf{x}^{(6)}) = \mathbf{0}$, then $t = 0.25$.

- $\mathbf{x}^{(6)} = \begin{bmatrix}0.25 & 0.25\end{bmatrix}^\top$.

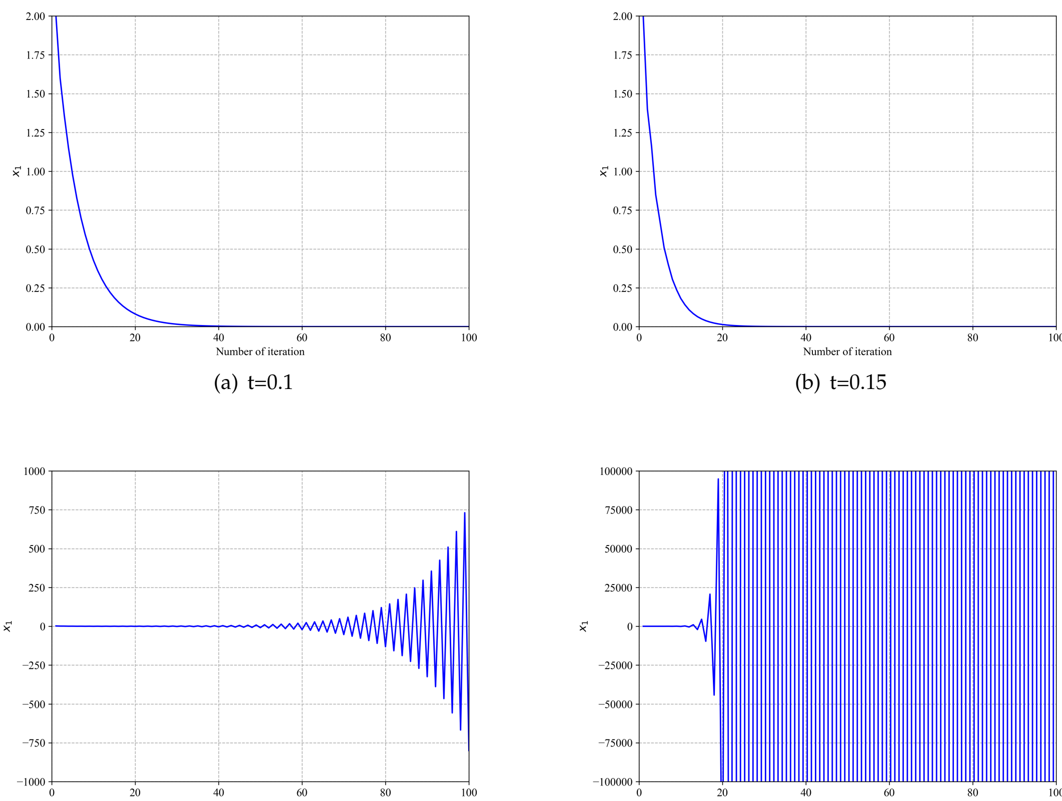
It can be summarized that $\mathbf{x}^{(2n)} = \begin{bmatrix}(\frac{1}{2})^{n-1} & (\frac{1}{2})^{n-1}\end{bmatrix}^\top$ from the previous iteration when $n \in \mathbb{N}^+$.

$$\lim_{n\to+\infty} \mathbf{x}^{(2n)} = \begin{bmatrix}0 & 0\end{bmatrix}^\top \tag{50}$$

Hence, $\mathbf{x}^* = \begin{bmatrix}0 & 0\end{bmatrix}^\top$ is the optimal solution to this problem.

(b) Figure(3) shows the convergence curve of different t settings. Figure(3(a)) and Figure(3(b)) show two convergence cases while Figure(3(c)) and Figure(3(d)) are two representative cases for non-convergence ones. $t = 0.2, 0.25, 0.30, ..., 1.0$ can not meet the requirements for convergence.

```python
import numpy as np
opx = []
# t value setting
t=0.1
# Initial guess
x= np.array([2,3])
matrix = np.array([[8,-4],[-4,4]])
for i in range(100):
i+=1
df = np.dot(matrix,x)
x1 = (x - df *t).copy()
dx = df*t
opx.append([i,x[0],x[1]])
x = x1.copy()
# Terminal condition
if(np.linalg.norm(dx)<=0.000001):
break
```



(a) t=0.1

(b) t=0.15

(c) t=0.2

(d) t=0.3

Figure 3: Convergence curve of different t settings

18. Consider the function

$$f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$$

Use the steepest descent method to minimize $f(x_1, x_2)$ with exact line search and with initial point $x^{(0)} = [4 \;\; 2 \;\; -1]^\top$. Perform three iteration and leave a comment with what you notice.

Once you are done, we want you to write a simple for loop that simulates steepest descent. How many iterations does it take to reach a solution with an error tolerance of 10^{-6}? The error tolerance is defined herein as

$$||x^{(k+1)} - x^{(k)}|| \leq \text{tol} = 10^{-6}$$

where $||\cdot||$ is the norm of the error between two successive iterations of the steepest descent method.

Solutions.

(a) First three iteration computed manually:.

$$f'(\mathbf{x}) = \begin{bmatrix} 4(x_1 - 4)^3 & 2(x_2 - 3) & 16(x_3 + 5)^3 \end{bmatrix}^\top \tag{51}$$

- 1^{st} iteration:

$$f'(\mathbf{x^{(0)}}) = \begin{bmatrix} 0 & -2 & 1024 \end{bmatrix}^\top \tag{52}$$

$$\mathbf{x^{(1)}} = \mathbf{x^{(0)}} - tf'(\mathbf{x^{(0)}}) = \begin{bmatrix} 4 & (2 + 2t) & (-1 - 1024t) \end{bmatrix}^\top \tag{53}$$

$$f(\mathbf{x^{(1)}}) = 4(1024t - 4)^4 + (2t + 2)^2 \tag{54}$$

$$t_1^* = \arg\min_{t>0} f(\mathbf{x^{(1)}})(t) = 0.00397 \tag{55}$$

$$\mathbf{x^{(1)}} = \begin{bmatrix} 4 & 2.008 & -5.065 \end{bmatrix} \tag{56}$$

- 2^{nd} iteration:

$$f'(\mathbf{x^{(2)}}) = \begin{bmatrix} 0 & -1.984 & -0.004394 \end{bmatrix}^\top \tag{57}$$

$$\mathbf{x^{(2)}} = \mathbf{x^{(1)}} - tf'(\mathbf{x^{(1)}}) = \begin{bmatrix} 4 & (2.008 + 1.984t) & (-5.065 + 0.004t) \end{bmatrix}^\top \tag{58}$$

$$f(\mathbf{x^{(2)}}) = 4(0.004t - 0.065)^4 + (1.984t - 0.992)^2 \tag{59}$$

$$t_2^* = \arg\min_{t>0} f(\mathbf{x^{(2)}})(t) = 0.50000 \tag{60}$$

$$\mathbf{x^{(2)}} = \begin{bmatrix} 4 & 3 & -5.067 \end{bmatrix} \tag{61}$$

- 3^{rd} iteration:

$$f'(\mathbf{x^{(2)}}) = \begin{bmatrix} 0 & 0 & 0.00481 \end{bmatrix}^\top \tag{62}$$

$$\mathbf{x^{(3)}} = \mathbf{x^{(2)}} - tf'(\mathbf{x^{(2)}}) = \begin{bmatrix} 4 & 3 & (-5.067 + 0.005t) \end{bmatrix}^\top \tag{63}$$

$$f(\mathbf{x^{(3)}}) = (0.005t - 0.067)^2 \tag{64}$$

$$t_3^* = \arg\min_{t>0} f(\mathbf{x^{(2)}})(t) = 13.4 \tag{65}$$

$$\mathbf{x^{(3)}} = \begin{bmatrix} 4 & 3 & -5 \end{bmatrix}^\top \tag{66}$$

$\mathbf{x^*} = \mathbf{x^{(3)}} = \begin{bmatrix} 4 & 3 & -5 \end{bmatrix}^\top$ is the optimal solution to the objective function.

(b) It takes the algorithm eight iterations to reach a solution with an error tolerance of 10^{-6}. The optimal solution is $\mathbf{x}^* = \begin{bmatrix} 4 & 3 & -5 \end{bmatrix}^\top$. Python-generated solution:

```
# Declare the dependencies
import numpy as np
# scipy.optimize is used to calculate the optimal t
from scipy.optimize import minimize_scalar
# Define a variable to contain the optima for each iteration
opx = []
# Initial guess
x= np.array([4,2,-1])
# Define the function of gradient
def gradient(x):
return np.array([4*(x[0]-4)**3 , 2*(x[1]-3) ,
16*(x[2]+5)**3])
# Define the objective function
def f(x):
return (x[0]-4)**4 + (x[1]-3)**2 + 4*(x[2]+5)**4
# the for loop
for i in range(100):
i+=1
df = gradient(x)
def ft(t,x=x):
return ((x[0] - (4*(x[0]-4)**3)*t)-4)**4 +
((x[1] - (2*(x[1]-3))*t)-3)**2 +
((x[2] - (16*(x[2]+5)**3)*t)+5)**4
res = minimize_scalar(ft, method='brent')
t = res.x
x1 = (x - df *t).copy()
dx = df*t
opx.append([i,x[0],x[1],x[2]])
x = x1.copy()
# Terminal condition
if(np.linalg.norm(dx)<=0.000001):
print(i)
print(x1)
# Quit the loop when the convergence condition reaches
break
```