The Chinese University of Hong Kong, Shenzhen
SDS · School of Data Science

Andre Milzarek · Fall Semester 2024/25

MDS 6106 — Optimization and Modeling

Exercise Sheet Nr.: Exercise 2

Name: Zijin CAI Student ID: 224040002

In the creation of this solution sheet, I worked together with:

Name: ______ Student ID: ______

Name: ______ Student ID: ______

Name: ______ Student ID: ______

For correction:

Exercise						$\sum$
Grading						

)72(b)equation.0.7)82(b)equation.0.8

(there may be some configurator errors, please ignore the content above)

Assignment A2.1 (Optimization Problem)

(a) the gradient of the function $f(x)$:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} \begin{pmatrix} 2x_1 + \frac{3}{2}x_1^2 - (2 + x_2^2) \\ -2x_1x_2 + \frac{1}{2}x_2^3 \end{pmatrix} \tag{1}$$

the Hessian matrix of the function $f(x)$:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 + 3x_1 & -2x_2 \\ -2x_2 & -2x_1 + \frac{3}{2}x_2^2 \end{bmatrix} \tag{2}$$

calculate all stationary points of the function $f(x)$ by solving the equation $\nabla f(x) = 0$:

$$\begin{bmatrix} 2x_1 + \frac{3}{2}x_1^2 - (2 + x_2^2) = 0 \\ -2x_1x_2 + \frac{1}{2}x_2^3 = 0 \end{bmatrix} \tag{3}$$

this gives two solution cases:

$$\begin{cases} x_2 = 0 \\ x_2^2 = 4x_1 \quad \text{or} \quad x_2 = \pm 2\sqrt{x_1} \end{cases} \tag{4}$$

for case 1 $x_2 = 0$:
we have $2x_1 + \frac{3}{2}x_1^2 - 2 = 0$, which gives $x_1 = -2$ or $x_1 = \frac{2}{3}$

for case 2: $x_2 = \pm 2\sqrt{x_1}$:
we have $2x_1 + \frac{3}{2}x_1^2 - 2 - 4x_1 = 0$, which gives $x_1 = 2$ or $x_1 = -\frac{2}{3}$
when $x_1 = 2$, we have $x_2 = \pm 2\sqrt{2}$
when $x_1 = -\frac{2}{3}$, we have $x_2 = \pm 2\sqrt{-\frac{2}{3}}$, which is not valid

hence, the stationary points x^* of the function $f(x)$ are: $(-2, 0)$, $(\frac{2}{3}, 0)$, $(2, 2\sqrt{2})$, and $(2, -2\sqrt{2})$

(b) for each stationary point x^*, calculate the Hessian matrix $\nabla^2 f(x^*)$ to determine the type of stationary point:

- if $\nabla^2 f(x^*)$ is positive definite, then x^* is a local minimizer
- if $\nabla^2 f(x^*)$ is negative definite, then x^* is a local maximizer
- if $\nabla^2 f(x^*)$ is indefinite, then x^* is a saddle point

for the stationary point $x^* = (-2, 0)$:
the Hessian matrix:

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 - 3 \times 2 & 0 \\ 0 & -2 \times (-2) + 0 \end{bmatrix} = \begin{bmatrix} -4 & 0 \\ 0 & 4 \end{bmatrix} \tag{5}$$

the determinant of the Hessian matrix is $-4 \times 4 - 0 \times 0 = -16$, and the trace is $-4 + 4 = 0$, which is indefinite, hence the stationary point $(-2, 0)$ is a saddle point

for the stationary point $x^* = \left(\frac{2}{3}, 0\right)$:
the Hessian matrix:

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 + 3 \times \left(\frac{2}{3}\right) & 0 \\ 0 & -2 \times \left(\frac{2}{3}\right) + 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & -\frac{4}{3} \end{bmatrix} \tag{6}$$

the determinant of the Hessian matrix is $4 \times -\frac{4}{3} - 0 \times 0 = -\frac{16}{3}$, the trace is $4 - \frac{4}{3} = \frac{8}{3}$, which is indefinite, hence the stationary point $\left(\frac{2}{3}, 0\right)$ is a saddle point

for the stationary point $x^* = \left(2, 2\sqrt{2}\right)$:
the Hessian matrix:

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 + 3 \times 2 & -4\sqrt{2} \\ -4\sqrt{2} & -4 + 12 \end{bmatrix} = \begin{bmatrix} 8 & -4\sqrt{2} \\ -4\sqrt{2} & 8 \end{bmatrix} \tag{7}$$

the determinant of the Hessian matrix is $8 \times 8 - 4 \times 4 \times 2 = 32$, the trace is $8 + 8 = 16$, which is positive definite hence the stationary point $(2, 2\sqrt{2})$ is a local minimizer

for the stationary point $x^* = \left(2, -2\sqrt{2}\right)$:
the Hessian matrix:

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 + 3 \times 2 & 4\sqrt{2} \\ 4\sqrt{2} & -4 + 12 \end{bmatrix} = \begin{bmatrix} 8 & 4\sqrt{2} \\ 4\sqrt{2} & 8 \end{bmatrix} \tag{8}$$

the determinant of the Hessian matrix is $8 \times 8 - 4 \times 4 \times 2 = 32$, the trace is $8 + 8 = 16$, which is positive definite hence the stationary point $(2, -2\sqrt{2})$ is a local minimizer

to conclude, the function $f(x)$ has two local minimizers at $(2, 2\sqrt{2})$ and $(2, -2\sqrt{2})$, and two saddle points at $(-2, 0)$ and $\left(\frac{2}{3}, 0\right)$

(c) consider the behavior of $f(x)$ as $||x|| \to \infty$, if $f(x)$ increases without bound as $||x|| \to \infty$, then one of the local minimum must be the global minimizer

if x_1 dominates x_2 (i.e., $|x_1| \gg |x_2|$):
then the dominant term of the function $f(x)$ is $\frac{1}{2}x_1^3$, which increases without bound as $x_1 \to \infty$ (to $-\infty$ as $x_1 \to -\infty$)

if x_2 dominates x_1 (i.e., $|x_2| \gg |x_1|$):
then the dominant term of the function $f(x)$ is $\frac{1}{4}x_2^4$, which increases without bound as $x_2 \to \infty$ (to $-\infty$ as $x_2 \to -\infty$)

however, if x_1 and x_2 grows with the similar magnitude, the behavior of the function $f(x)$ is more complicated, it can take arbitrary large positive or negative values depending on the direction of x tends to infinity

this implies the function $f(x)$ is not coercive, and the global minimizer may not be one of the local minimizers

here is some python code testing that, there exists function values smaller than the local minimum

[4]:
```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import minimize
```

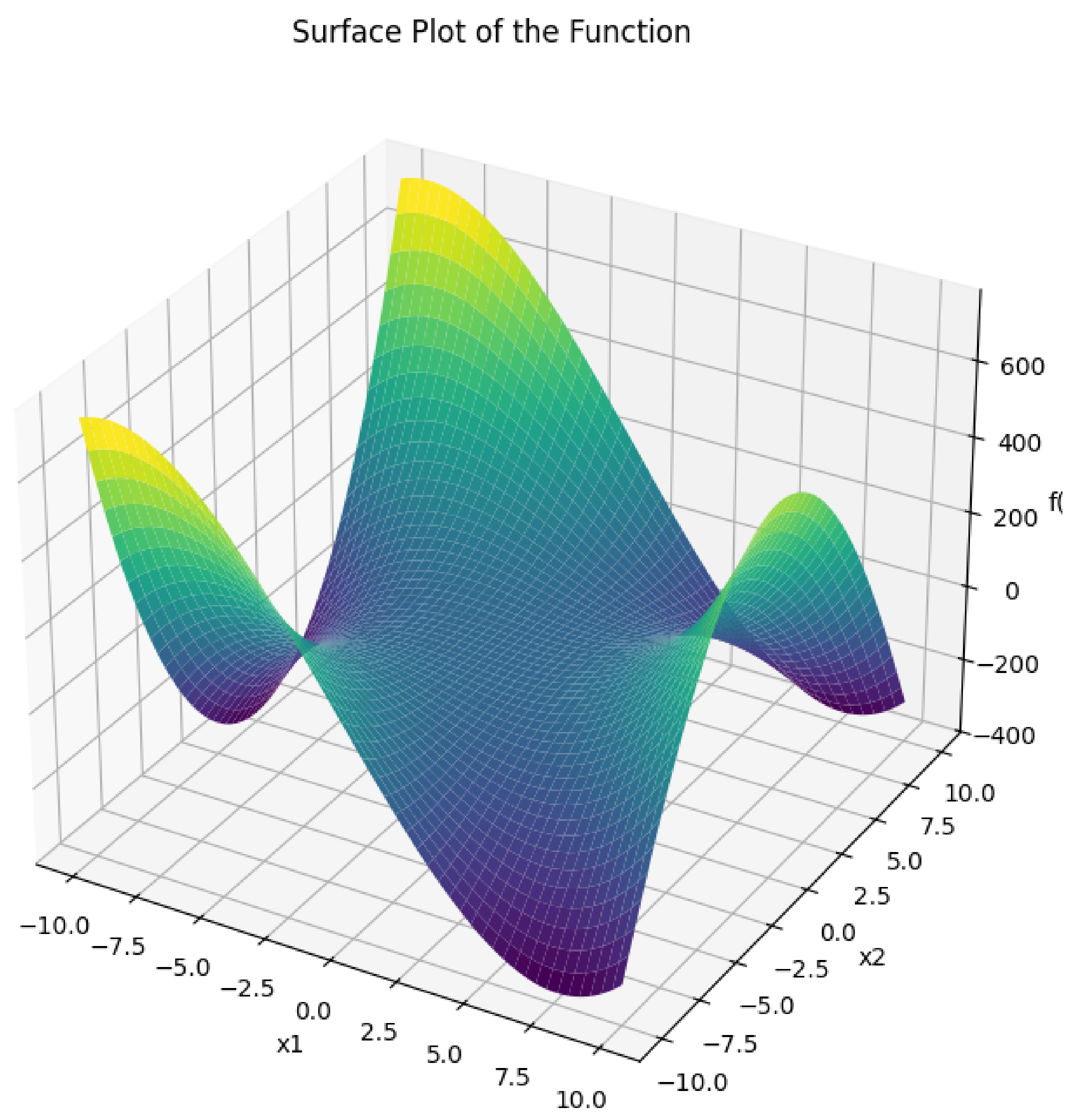
[5]:
```
def f(x):
    return x[0] ** 2 + 0.5 * x[0] ** 3 - x[0] * (2 + x[1] ** 2) + 0.0125 * x[1]␣
 ↪** 4
```

[6]:
```
x1 = np.linspace(-10, 10, 400)
x2 = np.linspace(-10, 10, 400)
X1, X2 = np.meshgrid(x1, x2)
Z = f([X1, X2])

fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(X1, X2, Z, cmap='viridis')

ax.set_title("Surface Plot of the Function")
ax.set_xlabel("x1")
ax.set_ylabel("x2")
ax.set_zlabel("f(x)")

plt.show()
```



[7]:
```
initial_guess = [0, 0]
result = minimize(f, initial_guess, method='BFGS')

print("Local minimum found at:", result.x)
print("Function value at local minimum:", result.fun)
```

```
Local minimum found at: [ 6.66664753e-01 -2.78457229e-26]
Function value at local minimum: -0.7407407407334132
```

[8]:
```
# Test points far from the origin
test_points = np.array([[10, 10], [-10, -10], [10, -10], [-10, 10]])
values = [f(point) for point in test_points]
```

```python
print("Function values at test points:", values)
```

```
Function values at test points: [-295.0, 745.0, -295.0, 745.0]
```

Assignment A2.2 (Coercivity)

(a) a cont. function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive if $\lim_{||x|| \to \infty} f(x) = +\infty$

$$f_1 : \mathbb{R}^2 \to \mathbb{R}, \quad \begin{aligned} f_1(x) &:= 2x_1^2 - 2x_1x_2^2 + x_2^2 + x_2^4 \\ &= (x_1 - x_2^2)^2 + x_1^2 + x_2^2 \end{aligned} \tag{9}$$

the suitable lower bound for $f_1(x)$ is $x_1^2 + x_2^2$, such that $f_1(x) \geq x_1^2 + x_2^2$
$lim_{||x|| \to \infty} f_1(x) \geq lim_{||x|| \to \infty} x_1^2 + x_2^2 = +\infty$, the lower bound $x_1^2 + x_2^2$ is coercive, and therefore f_1 is coercive as well

$$f_2 : \mathbb{R}^2 \to \mathbb{R}, \quad \begin{aligned} f_2(x) &:= 2x_1^2 - 8x_1x_2 + 7x_2^2 \\ &= 2(x_1 - 2x_2)^2 - x_2^2 \end{aligned} \tag{10}$$

choose the direction $x_1 = 2x_2$ such that: $x = (2x_2, x_2)$, then $f_2(x) = -x_2^2$, $lim_{||x|| \to \infty} f_2(x) = -\infty$ therefore, $f_2(x)$ is not coercive

$$f_3 : \mathbb{R}^n \to \mathbb{R}, \quad f_3(x) := \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{1 + ||\mathbf{x}||_2^2}, \text{ where} \mathbf{A} \in \mathbb{R}^{n \times n} \text{is symmetric and PD} \tag{11}$$

since $\mathbf{A}$ is symmetric and PD, all its eigenvalues are positive
let the eigenvalues of $\mathbf{A}$ be $\lambda_1, \lambda_2, \cdots, \lambda_n$
re-writing $\mathbf{x}^T \mathbf{A} \mathbf{x}$ as $\sum_{i=1}^n \lambda_i x_i^2$, then $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq \lambda_{\min} ||\mathbf{x}||_2^2$, thus:

$$\begin{aligned} f_3(x) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{1 + ||\mathbf{x}||_2^2} \quad &\geq \quad \frac{\lambda_{\min} ||\mathbf{x}||_2^2}{1 + ||\mathbf{x}||_2^2} \quad (\lambda_{\min} > 0) \\ &= \lambda_{\min} \times \frac{||\mathbf{x}||_2^2}{1 + ||\mathbf{x}||_2^2} \end{aligned} \tag{12}$$

because $\lim_{||x|| \to \infty} f_3(x) \geq \lim_{||x|| \to \infty} \lambda_{\min} \times \frac{||\mathbf{x}||_2^2}{1 + ||\mathbf{x}||_2^2} = \lambda_{\min}$,
as $\lim_{||x|| \to \infty} \frac{||\mathbf{x}||_2^2}{1 + ||\mathbf{x}||_2^2} = 1$, therefore $f_3(x)$ is not coercive

(b) the optimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}^2} f_1(x) = 2x_1^2 - 2x_1x_2^2 + x_2^2 + x_2^4 \tag{13}$$

the partial derivatives of f_1 with respect to x_1, x_2:

$$\nabla f_1(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} = 4x_1 - 2x_2^2 \\ \frac{\partial f_1}{\partial x_2} = -4x_1x_2 + 2x_2 + 4x_2^3 \end{pmatrix} \tag{14}$$

set these partial derivatives to zero, and solve for:

$$\begin{cases} 2x_1 = x_2^2 \\ 2x_2(x_2^2+1) = 0 \end{cases} \tag{15}$$

the solutions are:

$$\begin{cases} x_2 = 0 \\ x_2^2+1 = 0, \text{ which is not valid} \end{cases} \tag{16}$$

thus, the stationary point of f_1 is $(0,0)$

the Hessian matrix of f_1:

$$\nabla^2 f_1(x) = \begin{bmatrix} 4 & -4x_2 \\ -4x_2 & -4x_1+2+12x_2^2 \end{bmatrix} \quad \text{and} \quad \nabla^2 f_1(0,0) = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \tag{17}$$

which indicates that the Hessian matrix is positive definite at $(0,0)$, and therefore $(0,0)$ is a local minimizer of f_1

since $f_1 : \mathbb{R}^2 \to \mathbb{R}$ is coercive, $(x_1^*, x_2^*) = (0,0)$ is also the global minimizer of f_1
hence the optimization problem has a global solution of minimum at $(0,0)$, where $f_1(x) = 0$

Assignment A2.3 (Convex Sets)

(a) the convex set definition: a set $X \subseteq \mathbb{R}^n$ is convex if for any $\mathbf{x}, \mathbf{y} \in X$ and $\lambda \in [0,1]$, we have $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in X$

for any arbitrary $\mathbf{x}, \mathbf{y} \in X$ and $\lambda \in [0,1]$, such that :

$$\begin{aligned}
&\alpha \leq \mathbf{x}^3 \leq \beta \quad \text{and} \quad \alpha \leq \mathbf{y}^3 \leq \beta \\
&\text{need to show:} \quad \alpha \leq (\lambda\mathbf{x} + (1-\lambda)\mathbf{y})^3 \leq \beta \\
&\text{expand and re-write:} \quad \begin{aligned} &\lambda^3\mathbf{x}^3 + 3\lambda^2(1-\lambda)\mathbf{x}^2\mathbf{y} + 3\lambda(1-\lambda)^2\mathbf{x}\mathbf{y}^2 + (1-\lambda)^3\mathbf{y}^3 \\ &= \lambda^3\mathbf{x}^3 + (1-\lambda)^3\mathbf{y}^3 + 3\lambda^2(1-\lambda)(\mathbf{x}^2\mathbf{y} + \mathbf{x}\mathbf{y}^2) \end{aligned}
\end{aligned} \tag{18}$$

we know that $\alpha \leq \lambda^3\mathbf{x}^3 + (1-\lambda)^3\mathbf{y}^3 \leq \beta$
and $2\alpha \leq \mathbf{x}^2\mathbf{y} + \mathbf{x}\mathbf{y}^2 \leq \mathbf{x}^3 + \mathbf{y}^3 \leq 2\beta$, while $\lambda \in [0,1]$
we can solve for the upper bound for $3\lambda^2(1-\lambda)$ is $\frac{4}{9}$,where $\lambda = \frac{2}{3}$
thus, $\frac{8}{9}\alpha \leq 3\lambda^2(1-\lambda)(\mathbf{x}^2\mathbf{y} + \mathbf{x}\mathbf{y}^2) \leq \frac{8}{9}\beta$, which is also bounded by α and β

alternatively, the function $f(\mathbf{x})$ is convex on $\mathbb{R}^+$ because $f(\mathbf{x})'' = 6\mathbf{x}$ is non-negative for all $\mathbf{x} \in \mathbb{R}$, therefore the set $\{\mathbf{x} \in \mathbb{R} : f(\mathbf{x}) \leq \beta\}$ is convex for any $\beta \geq 0$, and the set $\{\mathbf{x} \in \mathbb{R} : \alpha \leq f(\mathbf{x})\}$ is also convex for any $\alpha \in \mathbb{R}$ regardless α is positive or negative
hence, the intersection of these two sets is also convex such that: $X = \left\{\mathbf{x} \in \mathbb{R} : \alpha \leq x^3 \leq \beta\right\}$ is convex

(b) consider two arbitrary points $\mathbf{x}, \mathbf{y}$, such that $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$, where $x_1, x_2, y_1, y_2 \geq 0$ and $x_1x_2 \geq 1, \quad y_1y_2 \geq 1$

for $\lambda \in [0,1]$, need to show: $(\lambda x_1 + (1-\lambda)y_1)(\lambda x_2 + (1-\lambda)y_2) \geq 1$

expand the left-hand side of the inequality:

$$\begin{aligned}
&(\lambda x_1 + (1-\lambda)y_1)(\lambda x_2 + (1-\lambda)y_2) \\
&= \lambda^2 x_1x_2 + (1-\lambda)^2 y_1y_2 + \lambda(1-\lambda)(x_1y_2 + x_2y_1)
\end{aligned} \tag{19}$$

since $x_1x_2 \geq 1$ and $y_1y_2 \geq 1$:

$$\begin{aligned}
&\lambda^2 x_1x_2 + (1-\lambda)^2 y_1y_2 \geq \lambda^2 + (1-\lambda)^2 \\
&\lambda(1-\lambda)(x_1y_2 + x_2y_1) \geq 2\lambda(1-\lambda)
\end{aligned} \tag{20}$$

adding these two inequalities:

$$(\lambda x_1 + (1-\lambda)y_1)(\lambda x_2 + (1-\lambda)y_2) \geq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) = 1 \geq 1 \tag{21}$$

therefore, the hyperbolic set $X := \left\{\mathbf{x} \in \mathbb{R}^2 : x_1x_2 \geq 1\right\}$, where $\mathbb{R}^2_+ := \left\{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \geq 0\right\}$, is convex

(c) for convex function $h : \mathbb{R}^n \to \mathbb{R}$
$h(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda h(\mathbf{x}) + (1-\lambda)h(\mathbf{y})$, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$

for convex set $X := \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) = 0\}$
$\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in X$, for any $\mathbf{x}, \mathbf{y} \in X$ and $\lambda \in [0, 1]$

the statement claims that the set X, which is defined by the level set where $h(x) = 0$, can only be convex if such the convex function $h(x)$ is a linear mapping

such statement is **FALSE**, because set X can be convex even if $h(x)$ is not a linear mapping

consider the convex function $h(x) = x^2 - 1$ (as $h''(x) = 2 > 0 \quad \forall x \in \mathbb{R}$)
then the solution set: $X = \left\{x \in \mathbb{R} : x^2 - 1 = 0\right\} = \{-1, 1\}$

consider any two points $x, y \in X$ and $\lambda \in [0, 1]$, such that x, y can only be either -1 or 1

and because any line segment between the two points lies entirely within the set, for example, we can choose λ either 0 or 1, then x and y is also in X, then set X is convex

(more general proof:)
suppose h is not a linear function, then there exist points $x, y \in X$, such that $h(\lambda x + (1-\lambda)y) \neq \lambda h(x) + (1-\lambda)h(y)$
suppose X is convex, this means we can find some $\lambda \in [0, 1]$, such that $\lambda x + (1-\lambda)y \in X$, where $h(\lambda x + (1-\lambda)y) = 0$

since $x, y \in X$, we have $h(x) = 0$ and $h(y) = 0$ such that:
$0 = h(\lambda x + (1-\lambda)y) \leq \lambda \times 0 + (1-\lambda) \times 0 = 0$
this inequality holds, but it does not imply that h must be a linear function

therefore, set X can be convex even if $h(x)$ is not a linear mapping

Assignment A2.4 (Convex Functions)

(a) $f : \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}$, $f(x_1, x_2) := \frac{x_1^2}{x_2}$, where $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$

the partial derivatives of f with respect to x_1, x_2:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{2x_1}{x_2} \\ -\frac{x_1^2}{x_2^2} \end{pmatrix} \tag{22}$$

the Hessian matrix of f:

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{bmatrix} \tag{23}$$

the determinant of the Hessian matrix is $\frac{4x_1^2}{x_2^4} - \frac{4x_1^2}{x_2^4} = 0$

the trace of the Hessian matrix is $\frac{2(x_1^2+x_2^2)}{x_2^3}$

as $x_1, x_2 \in \mathbb{R} : x > 0$, the trace $\frac{2(x_1^2+x_2^2)}{x_2^3}$ is non-negative, and the determinant is zero hence, the Hessian matrix is positive semi-definite
this implies that the function is convex over its specific domain

$f : \mathbb{R}^n \to \mathbb{R}, f(\mathbf{x}) := \frac{1}{2}||\mathbf{Ax} - \mathbf{b}||_2^2 + \mu||\mathbf{Lx}||_\infty$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{L} \in \mathbb{R}^{p\times n}$ and $\mu > 0$, and $||\cdot||_\infty$ denotes the maximum norm (i.e., $||\mathbf{x}||_\infty = \max_{i=1,\dots,n} |x_i|$ for all $\mathbf{x}$)

choose any arbitrary $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$
consider $z = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$, need to show:

$$\begin{aligned} f(z) &= \frac{1}{2}||\mathbf{Az} - \mathbf{b}||_2^2 + \mu||\mathbf{Lz}||_\infty \\ &= \frac{1}{2}||\mathbf{A}(\lambda\mathbf{x} + (\mathbf{1} - \lambda)\mathbf{y}) - \mathbf{b}||_2^2 + \mu||\mathbf{L}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})||_\infty \\ &\leq \lambda\left(\frac{1}{2}||\mathbf{Ax} - \mathbf{b}||_2^2 + \mu||\mathbf{Lx}||_\infty\right) + (1-\lambda)\left(\frac{1}{2}||\mathbf{Ay} - \mathbf{b}||_2^2 + \mu||\mathbf{Ly}||_\infty\right) \\ &= \lambda\frac{1}{2}||\mathbf{Ax} - \mathbf{b}||_2^2 + \lambda\mu||\mathbf{Lx}||_\infty + (1-\lambda)\frac{1}{2}||\mathbf{Ay} - \mathbf{b}||_2^2 + (1-\lambda)\mu||\mathbf{Ly}||_\infty \\ &= \lambda\frac{1}{2}||\mathbf{Ax} - \mathbf{b}||_2^2 + (1-\lambda)\frac{1}{2}||\mathbf{Ay} - \mathbf{b}||_2^2 + \mu||\mathbf{Ly}||_\infty \end{aligned} \tag{24}$$

follows the linearity of $\mathbf{A}$ and the convexity of the square norm, we can simplify the first term of the inequality that:

$$\frac{1}{2}||\mathbf{A}(\lambda\mathbf{x} + (\mathbf{1} - \lambda)\mathbf{y}) - \mathbf{b}||_2^2 \leq \lambda\frac{1}{2}||\mathbf{Ax} - \mathbf{b}||_2^2 + (1-\lambda)\frac{1}{2}||\mathbf{Ay} - \mathbf{b}||_2^2 \tag{25}$$

follows the linearity of $\mathbf{L}$ and the convexity of the maximum norm, we can simplify the second term of the inequality that:

$$\mu||\mathbf{L}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})||_\infty \leq \mu\left(\lambda||\mathbf{L}\mathbf{x}||_\infty + (1-\lambda)||\mathbf{L}\mathbf{y}||_\infty\right) \tag{26}$$

both the functions $\frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$ and $\mu||\mathbf{L}\mathbf{x}||_\infty$ are convex, therefore the joint function f is also convex over the domain

(b) for the convex mapping $f : \mathbb{R}^n \to \mathbb{R}_+$ and $g(\mathbf{x}) := (f(\mathbf{x}))^2$

choose any arbitrary $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$,
for $g(\lambda x + (1-\lambda)y) = (f(\lambda x + (1-\lambda)y))^2$, need to show:

$$(f(\lambda x + (1-\lambda)y))^2 \leq \lambda f(x)^2 + (1-\lambda)f(y)^2 \tag{27}$$

from the convexity of f, we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
then square both sides, we have:, $(f(\lambda x + (1-\lambda)y))^2 \leq (\lambda f(x) + (1-\lambda)f(y))^2$

expand the right-hand side of the inequality:

$$(\lambda f(x) + (1-\lambda)f(y))^2 = \lambda^2 f(x)^2 + (1-\lambda)^2 f(y)^2 + 2\lambda(1-\lambda)f(x)f(y) \tag{28}$$

hence, we have:

$$\begin{aligned} \lambda f(x)^2 + (1-\lambda)f(y)^2 &\leq \lambda^2 f(x)^2 + (1-\lambda)^2 f(y)^2 + 2\lambda(1-\lambda)f(x)f(y) \\ &= \lambda f(x)^2 + (1-\lambda)f(y)^2 + \lambda(1-\lambda)\left(f(x) - f(y)\right)^2 \end{aligned} \tag{29}$$

simplify the inequality, we have:

$$\lambda(1-\lambda)\left(f(x) - f(y)\right)^2 \geq 0 \tag{30}$$

since $\lambda, (1-\lambda) \in [0, 1]$, and $(f(x) - f(y))^2 \geq 0$ always holds, then the inequality $\lambda(1-\lambda)\left(f(x) - f(y)\right)^2 \geq 0$ is always true

therefore, the function $g(\mathbf{x}) = (f(\mathbf{x}))^2$ is convex as f is convex

for the convex mapping $f : \mathbb{R}^n \to \mathbb{R}_+$ and $g(\mathbf{x}) := \frac{1}{2}\left(||\mathbf{x}||^2 - 1\right)^2$

checking the Hessian of g can be very complex, instead, we can find a counterexample:

consider $x = (1,0)$ and $y = (0,1)$

$$\begin{aligned}
&g(x) = g(1,0) = \frac{1}{2}\left(1^2 - 1\right)^2 = 0 \\
&g(y) = g(0,1) = \frac{1}{2}\left(1^2 - 1\right)^2 = 0 \\
&g(x+y) = g(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}\left(||(\frac{1}{2}, \frac{1}{2})||^2 - 1\right)^2 = \frac{1}{8} \\
&\text{thus, we have:} \quad g(\frac{x+y}{2}) > \frac{g(x)+g(y)}{2}
\end{aligned} \tag{31}$$

which is contrary to the convexity of g, such that:

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y) \tag{32}$$

therefore the function $g(\mathbf{x}) = \frac{1}{2}\left(||\mathbf{x}||^2 - 1\right)^2$ is not convex

Assignment A2.5 (Optimal Location of a Warehouse)

(a) optimization problem to locate the warehouse

decision: choose the location of the warehouse $(x, y) \in \mathbb{R}^2$

objective: minimize the total transportation such that:

$$C = \sum_{i}^{4} (\mathbf{d}_i \times s_i \times unit\_cost\_per\_shipment) \quad \text{, where:} \tag{33}$$

C — the total transportation cost

$\mathbf{d}_i$ — the distance from the warehouse to the i-th customer is given by: $\mathbf{d}_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$

s_i — the shipment quantity to the i-th customer

$unit\_cost\_per\_shipment$ — 500 RMB per km

for $i \in \{\text{DJI}, \text{BYD}, \text{BGI}, \text{Huawei}\}$

constraint: the warehouse must be located within the industrial zone Z_1:

$$Z_1 = \left\{(x, y) \in \mathbb{R}^2 \colon -8 \leq x \leq 2, -5 \leq y \leq 3\right\} \tag{34}$$

hence, the optimization problem:

$$\min_{x,y} C(x, y) = \sum_{i}^{4} \left(\sqrt{(x - x_i)^2 + (y - y_i)^2} \times s_i\right) \times 500$$

$$\begin{aligned} \min_{x,y} C(x, y) = (&\sqrt{(x - 5)^2 + (y - 10)^2} \times 200 + \sqrt{(x - 10)^2 + (y - 5)^2} \times 150 \\ &+ \sqrt{(x - 0)^2 + (y - 12)^2} \times 200 + \sqrt{(x - 12)^2 + (y - 0)^2} \times 300) \times 500 \\ &\text{subject to} \qquad \text{-8} \leq x \leq 2, \quad -5 \leq y \leq 3 \qquad (35) \end{aligned}$$

(b) denote the location of the warehouse 1 as (x_1, y_1)

adding the building cost of warehouse 1 is $25,000,000$RMB, the optimization problem for the total cost of the warehouse 1 over 10 years:

$$\begin{aligned} \min_{x_1,y_1} C(x_1, y_1) = &\sum_{i}^{4} \left(\sqrt{(x_1 - x_i)^2 + (y_1 - y_i)^2} \times s_i\right) \times 500 \\ &\times 10 + 25,000,000 \end{aligned}$$

$$\min_{x_1,y_1} C(x_1,y_1) = (\sqrt{(x_1-5)^2+(y_1-10)^2} \times 200 + \sqrt{(x_1-10)^2+(y_1-5)^2} \times 150$$
$$+ \sqrt{(x_1-0)^2+(y_1-12)^2} \times 200 + \sqrt{(x_1-12)^2+(y_1-0)^2} \times 300)$$
$$\times 500 \times 10 + 25,000,000$$

$$\text{subject to} \qquad \text{-8} \le x_1 \le 2, \quad -5 \le y_1 \le 3 \qquad (36)$$

similarly, for choosing the location of the warehouse 2 as (x_2, y_2), such that: $Z_2 = \{(x_2, y_2) \in \mathbb{R}^2 \colon 7 \le x_2 \le 12, -3 \le y_2 \le 0\}$, the optimization:

$$\min_{x_2,y_2} C(x_2,y_2) = \sum_i^4 \left(\sqrt{(x_2-x_i)^2+(y_2-y_i)^2} \times s_i\right) \times 500$$
$$\times 10 + 40,000,000$$

$$\min_{x_2,y_2} C(x_2,y_2) = (\sqrt{(x_2-5)^2+(y_2-10)^2} \times 200 + \sqrt{(x_2-10)^2+(y_2-5)^2} \times 150$$
$$+ \sqrt{(x_2-0)^2+(y_2-12)^2} \times 200 + \sqrt{(x_2-12)^2+(y_2-0)^2} \times 300)$$
$$\times 500 \times 10 + 40,000,000$$

subject to $\qquad 7 \le x_2 \le 12, \quad -3 \le y_2 \le 0 \qquad (37)$

hence, the overall optimization problem:

$$\min_{(x_1,y_1),(x_2,y_2)} \{C(x_1,y_1), C(x_2,y_2)\}$$

$$\text{subject to} \qquad \begin{array}{l} -8 \le x_1 \le 2, \quad -5 \le y_1 \le 3 \\ \;\; 7 \le x_2 \le 12, \quad -3 \le y_2 \le 0 \end{array} \qquad (38)$$

(c) for the optimization in part (a):

$$\min_{x,y} C(x,y) = \sum_i^4 \left(\sqrt{(x-x_i)^2+(y-y_i)^2} \times s_i\right) \times 500$$

$$\text{subject to} \qquad \text{-8} \le x \le 2, \quad -5 \le y \le 3 \qquad (39)$$

this function is a sum of weighted Euclidean distances (i.e., $\left(\sqrt{(x-x_i)^2+(y-y_i)^2}\right)$), which is not a convex function, however, minimizing such the Euclidean distances is just as minimizing the square of the Euclidean distances, for instance, (i.e., $(x-x_i)^2+(y-y_i)^2$), which is then a convex function, and the sum of convex functions is also convex
therefore, the optimization objective function $C(x,y)$ is convex

the constraints are linear, which defines a convex set Z_1 (a rectangle in $\mathbb{R}^2$), hence the optimization problem is convex

for convex optimization problems, the local minimum is also the global minimum, therefore, the optimization problem has a global solution

for the optimization in part (b):

$$\min_{(x_1,y_1),(x_2,y_2)} \{C(x_1,y_1), C(x_2,y_2)\}$$

$$\text{subject to} \qquad \begin{aligned} -8 \le x_1 \le 2, \quad -5 \le y_1 \le 3 \\ 7 \le x_2 \le 12, \quad -3 \le y_2 \le 0 \end{aligned} \tag{40}$$

for each objective function $C(x_1,y_1)$ and $C(x_2,y_2)$, the components are still the sum of the convex function and a constant, (i.e., Euclidean distances and the building cost), hence each objective function is still convex

However, the piecewise nature of min $\{\}$ introduce non-convexity, due to the discrete choice between the two warehouses zones, although the constraints are also linear, which defines a convex set Z_1 and Z_2, hence the overall optimization problem is non-convex

despite the non-convexity, the optimization problem is still a mixed-integer linear programming (MILP) problem, which can be solved by the branch-and-bound algorithm and dual annealing

here is the python code example solving that the optimal solution is $(x_2, y_2) = (11.718, 0)$, and the total cost $C(x_2, y_2)$ is 42945486 RMB:

[1]:
```
import numpy as np
from scipy.optimize import dual_annealing

# Define the objective function for warehouse 1
def objective_warehouse1(x):
    x1, y1 = x
    return (np.sqrt((x1-5)**2 + (y1-10)**2) * 200 +
            np.sqrt((x1-10)**2 + (y1-5)**2) * 150 +
            np.sqrt((x1-0)**2 + (y1-12)**2) * 200 +
            np.sqrt((x1-12)**2 + (y1-0)**2) * 300) * 500 * 10 + 25000000

# Define the objective function for warehouse 2
def objective_warehouse2(x):
    x2, y2 = x
    return (np.sqrt((x2-5)**2 + (y2-10)**2) * 200 +
```

```python
            np.sqrt((x2-10)**2 + (y2-5)**2) * 150 +
            np.sqrt((x2-0)**2 + (y2-12)**2) * 200 +
            np.sqrt((x2-12)**2 + (y2-0)**2) * 300) * 500 * 10 + 40000000

# Combine the objective functions
def combined_objective(x):
    x1, y1, x2, y2 = x
    return objective_warehouse1((x1, y1)) + objective_warehouse2((x2, y2))

# Define the bounds for the variables
bounds = [(-8, 2), (-5, 3), (7, 12), (-3, 0)]

# Use dual_annealing to find the global minimum
result = dual_annealing(combined_objective, bounds)

# Print the results
print("Optimal solution:", result.x)
print("Minimum cost:", result.fun)
```

```
Optimal solution: [ 2.          3.         11.71795548  0.        ]
Minimum cost: 136888024.4294614
```