תרגיל בית 4 – מודלים לא ליניאריים בחקר ביצועים

מגישים :

אלעד בוכריס – 206202426

משה דידי – 311395834

שאלה 1 :

א.

```
def generic_grad(f, gf, lsearch, x0, eps):
    xk = x0
    xk_1 = xk - lsearch(f, xk, gf(xk)) * gf(xk)
    fs, gs, ts = [f(xk)], [np.linalg.norm(gf(xk))], [time.time()]
    while np.abs(f(xk) - f(xk_1)) > eps:
        xk = xk_1
        xk_1 = xk - lsearch(f, xk, gf(xk)) * gf(xk)
        fs.append(f(xk))
        gs.append(np.linalg.norm(gf(xk)))
        ts.append(time.time())
    return xk, fs, gs, ts
```

ב.

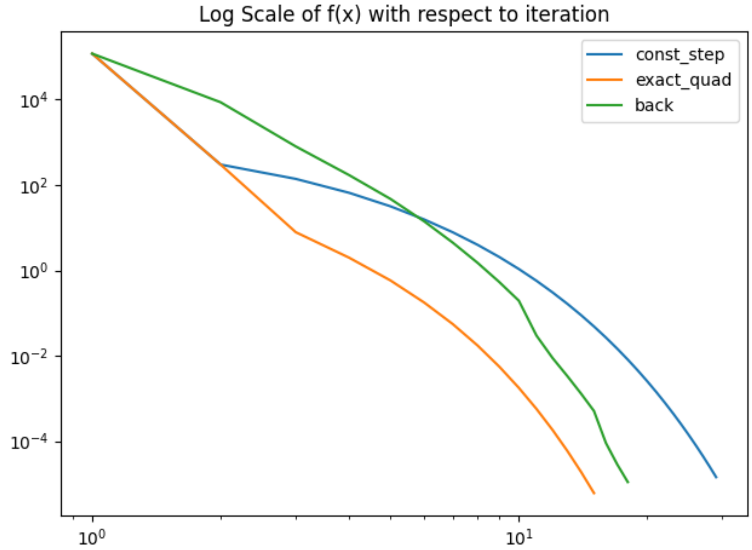
```
def const_step(s):
    return lambda f, xk, gk: s


def exact_quad(A):
    ATA = np.dot(A.T, A)
    np.linalg.cholesky(ATA)

    def lsearch(f, xk, gk):
        return 0.5 * np.square(np.linalg.norm(gk)) / np.square(np.linalg.norm(np.dot(A, gk)))

    return lsearch


def back(alpha, beta, s):
    def lsearch(f, xk, gk):
        t = s
        while f(xk - t * gk) >= f(xk) - alpha * t * np.square(np.linalg.norm(gk)):
            t *= beta
        return t

    return lsearch
```

ג.

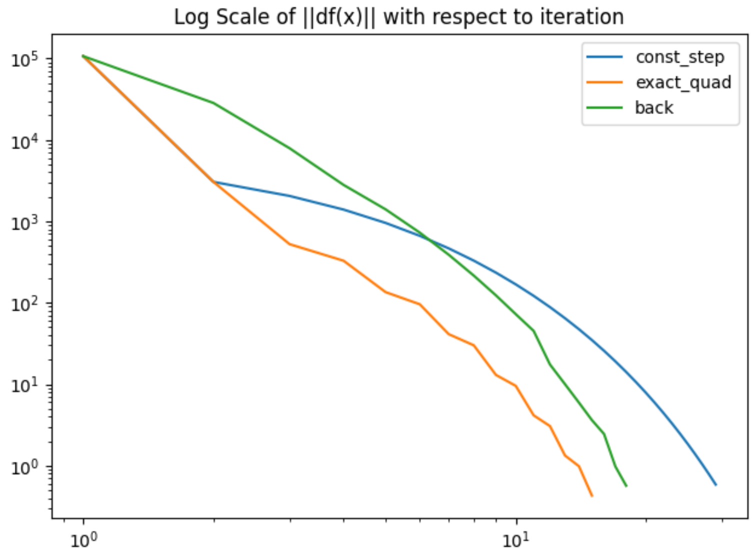
```
def f(x):
    A = np.array([[100, 2, 3, 4, 5],
                  [6, 100, 8, 9, 10],
                  [11, 12, 100, 14, 15],
                  [16, 17, 18, 100, 20],
                  [21, 22, 23, 24, 100]])
    return ((np.dot(x.T, A.T)).dot(A)).dot(x)


def gf(x):
    A = np.array([[100, 2, 3, 4, 5],
                  [6, 100, 8, 9, 10],
                  [11, 12, 100, 14, 15],
                  [16, 17, 18, 100, 20],
                  [21, 22, 23, 24, 100]])
    return 2 * ((np.dot(A.T, A)).dot(x))
```

```python
def generic_grad(f, gf, lsearch, x0, eps):
    xk = x0
    xk_1 = xk - lsearch(f, xk, gf(xk)) * gf(xk)
    fs, gs, ts = [f(xk)], [np.linalg.norm(gf(xk))], [time.time()]
    while np.abs(f(xk) - f(xk_1)) > eps:
        xk = xk_1
        xk_1 = xk - lsearch(f, xk, gf(xk)) * gf(xk)
        fs.append(f(xk))
        gs.append(np.linalg.norm(gf(xk)))
        ts.append(time.time())
    return xk, fs, gs, ts


def q1():
    A = np.array([[100, 2, 3, 4, 5],
                  [6, 100, 8, 9, 10],
                  [11, 12, 100, 14, 15],
                  [16, 17, 18, 100, 20],
                  [21, 22, 23, 24, 100]])
    s = 1 / (2 * np.max(np.linalg.eigvals(np.dot(A.T, A))))
    x0 = np.array([1, 1, 1, 1, 1])
    eps = 1 / np.power(10, 5)

    _const = generic_grad(f, gf, const_step(s), x0, eps)
    _exact = generic_grad(f, gf, exact_quad(A), x0, eps)
    _back = generic_grad(f, gf, back(0.5, 0.5, 1), x0, eps)
    plt.loglog(np.arange(1, len(_const[1]) + 1), _const[1], label="const_step")
    plt.loglog(np.arange(1, len(_exact[1]) + 1), _exact[1], label="exact_quad")
    plt.loglog(np.arange(1, len(_back[1]) + 1), _back[1], label="back")
    plt.title("Log Scale of f(x) with respect to iteration")
    plt.legend()
    plt.show()

    plt.loglog(np.arange(1, len(_const[2]) + 1), _const[2], label="const_step")
    plt.loglog(np.arange(1, len(_exact[2]) + 1), _exact[2], label="exact_quad")
    plt.loglog(np.arange(1, len(_back[2]) + 1), _back[2], label="back")
    plt.title("Log Scale of ||df(x)|| with respect to iteration")
    plt.legend()
    plt.show()

    plt.semilogy(_const[3], _const[2], label="const_step")
    plt.semilogy(_exact[3], _exact[2], label="exact_quad")
    plt.semilogy(_back[3], _back[2], label="back")

    plt.title("||df(x)|| with respect to time")
    plt.legend()
    plt.show()
```

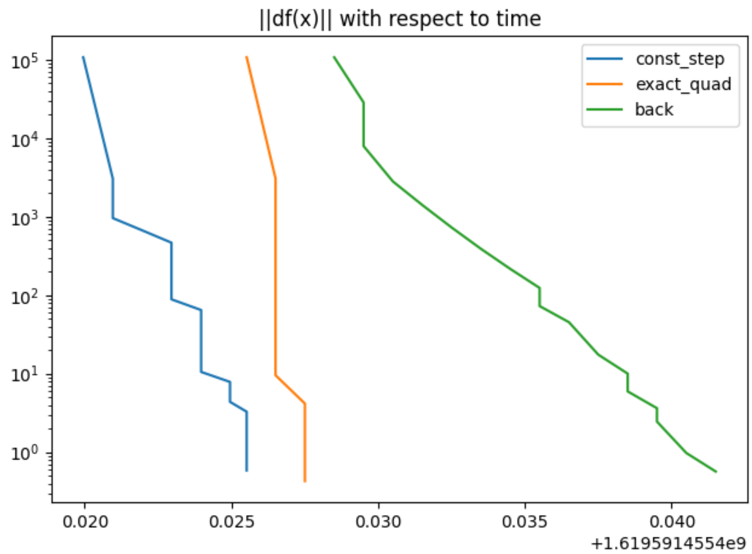
תרשימים:



Log Scale of f(x) with respect to iteration
const_step
exact_quad
back
10^4
10^2
10^0
10^{-2}
10^{-4}
10^0
10^1



Log Scale of ||df(x)|| with respect to iteration
const_step
exact_quad
back
10^5
10^4
10^3
10^2
10^1
10^0
10^0
10^1



||df(x)|| with respect to time
const_step
exact_quad
back
10^5
10^4
10^3
10^2
10^1
10^0
0.020
0.025
0.030
0.035
0.040
+1.6195914554e9

שאלה 2 :

א. כאשר :

D_{ij}- המרחק בין אובייקט i לאובייקט j.

נשים לב כי בעיית האופטימיזציה מוצאת ערכי x כך שלכל $i \neq j$ מתקיים

$$\left\|x_i - x_j\right\|^2 \approx D_{ij}^2$$

כלומר, בעיית האופטימיזציה מנסה למצוא $\left\|x_i - x_j\right\|$ אשר קרובה ככל הניתן ל- D_{ij}.

ולכן היא מוצאת קונפיגורציה עבור כל אובייקט.

ב. ראשית, ניתן לראות כי $\left\|x_i - x_j\right\|^2 - D_{ij}^2$ גזיר ולכן $S(X)$ גזירה כסכום של פונקציות גזירות.

בתרגול, ראינו כי עבור כל פונקציה חסומה מלרע f שיטת הגרדיאנט מתכנסת. ניתן לראות כי לאור העובדה ש - $S(X)$ הינה סכום ריבועים מתקיים :

$$S(X) \geq 0 \ \forall X$$

כלומר, חסומה מלרע ולכן סדרת הערכים $\{S(X^k)\}_{k\geq 0}$ מתכנסת.

ג. נגדיר :

$$u(x) = \left\|x_i - x_j\right\|^2 - D_{ij}^2$$

ו-

$$S(u) = \sum_{j=1}^{N} (u)^2$$

$$S\big(u(x)\big) = \sum_{j=1}^{N} \left(\left\|x_i - x_j\right\|^2 - D_{ij}^2\right)^2 \quad \forall i \in [N]$$

$$\frac{\partial u}{\partial x_i} = 2\big(x_1 - x_j\big)$$

$$\frac{\partial S}{\partial u} = \sum_{(j=1)}^{N} u'(x) \cdot 2u = 4\sum_{j=1}^{n} \big(x_i - x_j\big)\left(\left\|x_i - x_j\right\|^2 - D_{ij}^2\right) \ \forall i \in [N]$$

ולכן :

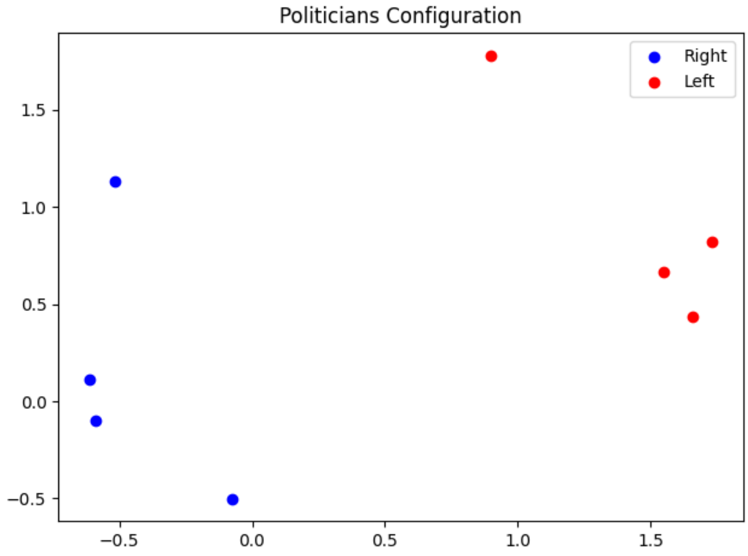
$$\nabla S(X) = 4\begin{pmatrix} \sum_{j=1}^{N} \big(x_1 - x_j\big)\left(\left\|x_1 - x_j\right\|^2 - D_{1j}^2\right) \\ \vdots \\ \sum_{j=1}^{N} \big(x_N - x_j\big)\left(\left\|x_N - x_j\right\|^2 - D_{Nj}^2\right) \end{pmatrix}$$

ד. בהינתן קונפיגורציה

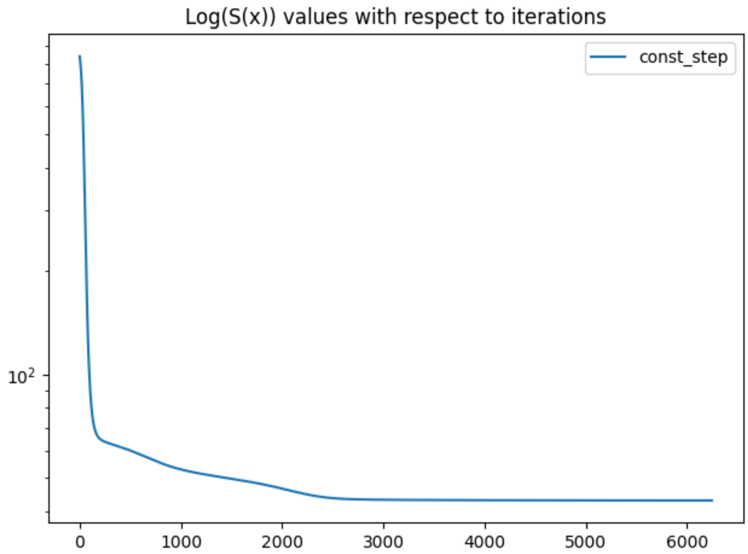
$$x_1 = \cdots = x_N = \begin{pmatrix} c \\ c \end{pmatrix} \to \big(x_i - x_j\big) = 0 \ \forall j, i \in [N] \to \nabla S(X) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

אם נאתחל את שיטת הגרדיאנט עם קונפיגורציה זו אזי כל צעד שווה לקודמו ולכן השיטה תתכנס תמיד לאחר איטרציה אחת לאותה נקודה התחלתית. המחשה : $x_{k+1} = x_k - t \cdot 0$.

ה.   קופיגורציית חברי הכנסת:



ניתן לראות כי קיים הבדל בין מפלגת הימין למפלגת השמאל.



השיטה התכנסה לאחר סדר גודל של 1000 איטרציות.

קוד שאלה 2:

```
def S(X):
    O = np.array([[1, 0, 0, -1, -1, -1, -1, -1, 0],
                  [0, 1, -1, -1, -1, -1, -1, -1, -1],
                  [1, 1, -1, -1, -1, -1, -1, -1, -1],
                  [0, 0, -1, 1, -1, -1, -1, -1, -1],
                  [-1, -1, -1, 1, 1, 1, 1, 1, 0],
                  [-1, -1, 1, 1, 1, 0, 1, 0, 1],
                  [-1, -1, 1, 1, 1, 1, 1, 1, 1],
                  [-1, -1, 1, 1, 0, 1, 1, 0, 0]])
    res = 0
    for i in range(len(X)):
        for j in range(len(X)):
            res += np.square(np.square(np.linalg.norm(X[i, :] - X[j, :])) - D(O[i, :], O[j, :]))
    return res
```

```
def gS(X):
    O = np.array([[1, 0, 0, -1, -1, -1, -1, -1, 0],
                  [0, 1, -1, -1, -1, -1, -1, -1, -1],
                  [1, 1, -1, -1, -1, -1, -1, -1, -1],
                  [0, 0, -1, 1, -1, -1, -1, -1, -1],
                  [-1, -1, -1, 1, 1, 1, 1, 1, 0],
                  [-1, -1, 1, 1, 1, 0, 1, 0, 1],
                  [-1, -1, 1, 1, 1, 1, 1, 1, 1],
                  [-1, -1, 1, 1, 0, 1, 1, 0, 0]])
    res = []
    for i in range(len(X)):
        res.append(
            sum([(X[i, :] - X[j, :]) *
                 (np.square(np.linalg.norm(X[i, :] - X[j, :])) - D(O[i, :], O[j, :])) for j in
range(len(X))]))
    return np.array(res)


def D(o1, o2):
    return np.linalg.norm(o1 - o2)


def q2():
    s = 1 / 1000
    X = np.random.rand(8, 2)
    eps = 1 / np.power(10, 5)

    _const = generic_grad(S, gS, const_step(s), X, eps)
    plt.scatter(_const[0][:, 0][:4], _const[0][:, 1][:4], label="Right", color='blue')
    plt.scatter(_const[0][:, 0][4:], _const[0][:, 1][4:], label="Left", color='red')
    plt.title("Politicians Configuration")
    plt.legend()
    plt.show()

    plt.semilogy(np.arange(len(_const[1])), _const[1], label="const_step")
    plt.title("Log(S(x)) values with respect to iterations")
    plt.legend()
    plt.show()
```

שאלה 3 :

תהי $f\colon \mathbb{R}^2 \to \mathbb{R}$ המוגדרת על־ידי

$$.f\left(x,y\right)=x^2+y^4-y^2$$

א. נקודות סטציונריות :

$$\nabla f(x,y) = (2x, 4y^3 - 2y) \to (0,0), \left(0, \pm\frac{1}{\sqrt{2}}\right)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & (12y^2 - 2) \end{pmatrix}$$

עבור $(0,0)$ נקודת אוכף מכיוון שההסיאן לא מוגדר.

עבור $\left(0, \pm\frac{1}{\sqrt{2}}\right)$ נקודות מינימום לוקלי ממש, מכיוון שההסיאן מוגדר חיובית.

ננסה לחסום את הפונקציה מלמטה :

$$f\left(0, \frac{1}{\sqrt{2}}\right) = -\frac{1}{4}$$

ננסה לחסום על ידי ערך זה

$$x^2 + y^4 - y^2 = x^2 + \left(y^2 - \frac{1}{2}\right)^2 - \frac{1}{4} \geq -\frac{1}{4}$$

ולכן פונקציה זו חסומה מלמטה ונקודות אלו הן מינימום גלובלי.

ב. האיבר הכללי נתון על ידי :

$$x^{k+1} = x^k - t^k \nabla \mathrm{f}(\mathrm{x}^\mathrm{k}) = (x^k, 0) - t^k(2x^k, 0) = (x^k(1 - 2t^k), 0)$$

מכיוון שידוע כי הסדרה

$f\{(x^k, y^k)\}_{k\geq 0}$ מונוטונית יורדת ממש, אזי כאשר k שואף לאינסוף ערכי הפונקציה שואפים לנקודה מינימום של $f(x,0)$ לאור נקודת ההתחלה בה $\nabla f(x^0, y^0) = (2x^0, y^0)$ נקודת המינימום של החתך הזה של הפונקציה מתקבלת בנקודה $(0,0)$.

לכן, **לא** נוכל להגיע לנקודות $\left(0, \pm\frac{1}{\sqrt{2}}\right)$ ונהיה חייבים להתכנס לנקודה $(0,0)$.

ג. פלטים :

x: [0.00015325 0.   ]

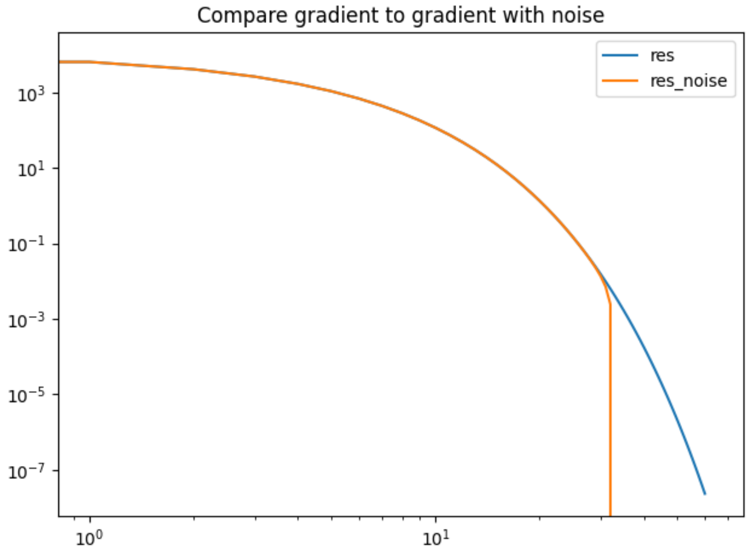
x_noise: [7.84642227e-05 7.07061296e-01]

f(x): [10000, 6400.0, 4096.0, 2621.4400000000005, 1677.7216, 1073.741824, 687.19476736, 439.8046511104001, 281.4749767106561, 180.14398509481993, 115.29215046068474, 73.78697629483824, 47.22366482869648, 30.223145490365745, 19.342813113834072, 12.379400392853807, 7.922816251426436, 5.070602400912918, 3.245185536584268, 2.0769187434139313, 1.329227995784916, 0.8507059173023463, 0.5444517870735016, 0.34844914372704106, 0.22300745198530625, 0.14272476927059602, 0.09134385233318146, 0.05846006549323614, 0.03741444191567113, 0.023945242826029525, 0.015324955408658894, 0.009807971461541692, 0.006277101735386684, 0.0040173451106474784, 0.002571100870814386, 0.0016455045573212073, 0.0010531229166855726, 0.0006739986666787665, 0.0004313591466744106, 0.0002760698538716228, 0.0001766847064778386, 0.00011307821214581669, 7.237005577332267e-05, 4.631683569492652e-05, 2.9642774844752967e-05, 1.8971375900641897e-05, 1.2141680576410814e-05, 7.770675568902921e-06, 4.973232364097869e-06, 3.182868713022636e-06, 2.037035976334487e-06, 1.3037030248540719e-06, 8.343699359066062e-07, 5.339967589802278e-07, 3.4175792574734585e-07, 2.1872507247830136e-07, 1.3998404638611285e-07, 8.958978968711223e-08, 5.733746539975183e-08, 3.669597785584117e-08, 2.348542582773835e-08]

f(x_noise): [10000, 6400.073569519582, 4096.047084454117, 2621.4701339953167, 1677.7408856773445, 1073.7541667187927, 687.2026665348478, 439.80970634444424, 281.4782117179285, 180.1460550062519, 115.29347449376193, 73.78782265326487, 47.22420502534368, 30.223489095473685, 19.343029967244586, 12.379534781513485, 7.9228959278044675, 5.070644275333603, 3.2451992059261063, 2.07690858479429, 1.329194269466719, 0.850645131390251, 0.5443564400370643, 0.3483068535613863, 0.2227993838478841 7, 0.14242317103610988, 0.09090838489551309, 0.05783245604535682, 0.03651077581950359, 0.02264492236386943, 0.013454969969783716, 0.007120626844719858, 0.002418750074308998, -0.0015149991898729962, -0.005346691023524981, -0.009656101055428717, -0.015017270596257909, -0.02205468902152854, -0.03147147132288892, -0.044032898195211785, -0.060474263053871574, -0.0812946697358274, -0.10642138987267839, -0.13481838111837474, -0.16427497547706144, -0.19172776950434794, -0.21426451755796067, -0.23033483497172427, -0.24024994280665773, -0.24558776459904794, -0.24814364621937834, -0.24925967472957977, -0.24971536845491116,

-0.2498931107674762, -0.2499604146340919, -0.24998544007252876, -0.24999465011242814, -0.24999802455075618, -0.24999926151274127, -0.24999971760387227, -0.2499998879571338, -0.24999995306579847, -0.2499999788901266, -0.24999998970585385]

גרף:



השיטה המורעשת התכנסה למינימום הגלובלי של f בנקודה $\left(0,\frac{1}{\sqrt{2}}\right)$. אנו צפינו זאת, מכיוון שהוספת β "הוציאה" את השיטה מהחתך $(x,0)$ ואפשרה לה להתקדם לכיוון המינימום הגלובלי של הפונקציה.
חסרון של השיטה המורעשת יכול להתבטא במספר רב יותר של איטרציות, מכיוון שאנו לא הולכים בדיוק נגד כיוון הגרדיאנט, אשר מביא לירידה מקסימלית. לא יכולנו לדעת לאיזו נקודה תתכנס השיטה, כי באופן דומה היא יכלה להתכנס לנקודה $\left(0,-\frac{1}{\sqrt{2}}\right)$ כפי שראינו בהרצות אחרות של הפונקציה.

קוד שאלה 3:

```
def f_q3(x):
    return np.square(x[0]) + np.power(x[1], 4) - np.square(x[1])

def df_q3(x):
    return np.array([2 * x[0], 4 * np.power(x[1], 3) - 2 * x[1]])

def generic_grad_noise(f, gf, lsearch, x0, eps, mu, sigma):
    xk = x0
    xk_1 = xk - lsearch(f, xk, gf(xk)) * gf(xk) + np.random.normal(loc=mu, scale=sigma,
size=(len(xk)))
    fs, gs, ts = [f(xk)], [np.linalg.norm(gf(xk))], [time.time()]
    while np.abs(f(xk) - f(xk_1)) > eps:
        xk = xk_1
        xk_1 = xk - lsearch(f, xk, gf(xk)) * gf(xk)
        fs.append(f(xk))
        gs.append(np.linalg.norm(gf(xk)))
        ts.append(time.time())
    return xk, fs, gs, ts

def ex3(mu, sigma, x0, epsilon):
    x0 = np.array(x0)
    res = generic_grad(f_q3, df_q3, const_step(1 / 10), x0, epsilon)
    res_noise = generic_grad_noise(f_q3, df_q3, const_step(1 / 10), x0, epsilon, mu, sigma)
    plt.loglog(np.arange(len(res[1])), res[1], label="res")
    plt.loglog(np.arange(len(res_noise[1])), res_noise[1], label="res_noise")
    plt.title("Compare gradient to gradient with noise")
    plt.legend()
    plt.show()
    return res[0], res_noise[0], res[1], res_noise[1]

def q3():
    res = ex3(0, 0.0005, [100, 0], 1 / np.power(10, 8))
    print(f'x: {res[0]}\nx_noise: {res[1]}\nf(x): {res[2]}\nf(x_noise): {res[3]}\n')
```

שאלה 4 :

א. יש להראות כי איטרציית הגראדינט שקולה לפתרון הבעיה :

$$.\mathbf{x}^{k+1} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ g_k(\mathbf{x}) \equiv f\left(\mathbf{x}^k\right) + \nabla f\left(\mathbf{x}^k\right)^T \left(\mathbf{x} - \mathbf{x}^k\right) + \frac{1}{2t^k} \left\|\mathbf{x} - \mathbf{x}^k\right\|_2^2 \right\}$$

$$\nabla g = \nabla \mathrm{f}\left(\mathrm{x}^{\mathrm{k}}\right)^T + \frac{1}{t^k}\left(x - x^k\right) \stackrel{!}{=} 0_n \rightarrow$$

$$x = x^k - t^k \nabla \mathrm{f}\left(\mathrm{x}^{\mathrm{k}}\right)$$

לכן, זוהי נקודה סטציונרית.

$$\nabla^2 g = diag\left(\frac{1}{t^k}\right) \stackrel{1}{\succ} 0$$

כלומר, זוהי נקודת המינימום של הפונקציה. ולפי איטרציית הגרדיאנט היא שווה ל- x^{k+1}.

1. באיטרצית הגרדיאנט מתקיים כי $t^k > 0$ ולכן ההסיאן מוגדר חיובית.

ב. יש להוכיח כי

$$\nabla f\left(x^k\right) \neq 0, t^k > 0, f(x) \leq g(x), \forall x \in \mathbb{R}^n \rightarrow f\left(x^{k+1}\right) < f\left(x^k\right)$$

$$g\left(x^{k+1}\right) \leq g(x), \forall\, x \in \mathbb{R}^n$$

ידוע כי

$$g\left(x^{k+1}\right) = f\left(x^k\right) + \nabla f\left(x^k\right)\left(x^k - t^k \nabla \mathrm{f}\left(\mathrm{x}^{\mathrm{k}}\right) - x^k\right) + \frac{1}{2t^k}\left\|x^k - t^k \nabla \mathrm{f}\left(\mathrm{x}^{\mathrm{k}}\right) - x^k\right\| =$$

$$f\left(x^k\right) - \frac{t^k}{2}\left\|\nabla f\left(x^k\right)\right\|^2$$

נתון כי

$$\nabla f\left(x^k\right) \neq 0, t^k > 0$$

ולכן מתקיים :

$$f\left(x^k\right) - \frac{t^k}{2}\left\|\nabla f\left(x^k\right)\right\|^2 < f\left(x^k\right)$$

נשים לב כי מתקיים

$$f\left(x^{k+1}\right) \leq g\left(x^{k+1}\right) = f\left(x^k\right) - \frac{t^k}{2}\left\|\nabla f\left(x^k\right)\right\|^2 < f\left(x^k\right)$$

מש"ל.