תרגיל בית 6 – מודלים לא ליניאריים בחקר ביצועים

מגישים :

אלעד בוכריס – 206202426

משה דידי – 311395834

שאלה 1 :

נתונה הפונקציה :

$$\min_{x\in interior(\mathcal{P})}\{f(x) \equiv -\sum_{i=1}^{m}\ln(b_i - a_i^T x) : a_i^T \in \mathbb{R}^n \neq 0_n, b \in \mathbb{R}^m, \mathcal{P} = \{x \in \mathbb{R}^n : Ax \leq b\}$$

כאשר הפונקציה שאנו רוצים למזער מוגדרת להיות המרכז האנליטי של $\mathcal{P}$.

א. נוכיח כי $\nabla^2 f(x) \succcurlyeq 0$ עבור x השייכים לפוליגון והפונקציה מוגדרת בהם.

נגדיר :

$$y_i = b_i - a_i^T x$$

ונגדיר :

$$g(y) = -\sum_{i=1}^{m}\ln(y_i)$$

ניתן לראות כי

$$f(x) = g(Bx + b)$$

כאשר :

$$b = b\ (defined\ in\ question)$$
$$B = -A\ (defined\ in\ question)$$

לכן :

$$\nabla^2 f(x) = B^T \nabla^2 g(y) B = A^T \nabla^2 g(y) A$$

נחשב את :

$$\nabla g(y) = \begin{pmatrix} -\frac{1}{y_1} \\ -\frac{1}{y_2} \\ \vdots \\ -\frac{1}{y_m} \end{pmatrix}$$

$$\nabla^2 g(y) = diag\left(\frac{1}{y_i^2}\right)_{i=1}^{m}$$

ניתן לראות כי

$$\nabla^2 g(y) \overset{1}{\succcurlyeq} 0$$

1. מטריצה אלכסונית שכל איבר באלכסון חיובי.

לכן, אנו יודעים כי :

$$u^T \nabla^2 g(y) u > 0 : u \neq 0_m \to u^T \nabla^2 f(x) u = u^T A \nabla^2 g(y) A^T u \overset{2}{\geq} 0$$

2. מכיוון שייתכן ו-$A^T u = 0$ אז הוא לא מוגדר חיובית.

ב. נניח כי

$$A = (-1) \in \mathbb{R}^{1\times 1}, b = -1 \in \mathbb{R}^1$$

נשים לב כי

$$x \in interior(\mathcal{P}) \to x \geq 1$$

$$f(x) = -\ln(-1 + x) : x > 1 \to f'(x) = -\frac{1}{-1 + x} < 0\ \forall x \in interior(\mathcal{P})$$

ולכן הפונקציה יורדת לכל x בתחום, כלומר, אין לה מינימום גלובלי או לוקלי ולכן אין פתרון מיטבי לבעיה.

ג. נתונים :

$$A = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$$
$$b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

ראשית, נמצא את התחום שמקיים את האינטריור :

$$-x - y \le 0 \to x + y \ge 0$$
$$x + y \le 1$$

ולכן :

$$0 \le x + y \le 1$$

נכתוב את

$$\nabla f(x) = -A^T \nabla g(Ax + b) = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -\frac{1}{x+y} \\ -\frac{1}{1-x-y} \end{pmatrix} = \begin{pmatrix} -\frac{1}{x+y} + \frac{1}{1-x-y} \\ \frac{1}{1-x-y} - \frac{1}{x+y} \end{pmatrix}$$

לכן נקבל כי בנקודה סטציונרית מתקיים :

$$x + y = (1 - x - y) \to 2x + 2y = 1 \to y = \frac{1}{2} - x$$

ניתן לראות כי ישר זה שייך ל-*interior of P*.

נכתוב את :

$$\nabla f(x) = A^T \nabla^2 g(y) A = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{(x+y)^2} & 0 \\ 0 & \frac{1}{\left(1-(x+y)\right)^2} \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} -\frac{1}{(x+y)^2} & \frac{1}{\left(1-(x+y)\right)^2} \\ -\frac{1}{(x+y)^2} & \frac{1}{\left(1-(x+y)\right)^2} \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{(x+y)^2} + \frac{1}{\left(1-(x+y)\right)^2} & \left(\frac{1}{(x+y)^2} + \frac{1}{\left(1-(x+y)\right)^2}\right) \\ +\left(\frac{1}{(x+y)^2} + \frac{1}{\left(1-(x+y)\right)^2}\right) & \frac{1}{(x+y)^2} + \frac{1}{\left(1-(x+y)\right)^2} \end{pmatrix}$$

נסמן: $a = \frac{1}{(x+y)^2} + \frac{1}{\left(1-(x+y)\right)^2}$

$$\begin{pmatrix} a & a \\ a & a \end{pmatrix} - \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix} = \begin{pmatrix} a-m & a \\ a & a-m \end{pmatrix}$$
$$trace = 2(a - m)$$

ראשית, עבור $m > a$ מתקבלים ע״ע שליליים ולכן המטריצה לא יכולה להיות מוגדרת אי שלילית.

$$\det \quad = (a-m)^2 - a^2 = -2am + m^2 = m(m - 2a)$$

נחלק למקרים :

$$m(m - 2a) > 0 \to m > 2a$$

על מנת שגם ה-*trace* יהיו חיובים, צריך להתקיים $m < a$ ולכן התחום ריק.

$$m(m - 2a) < 0 \to cant\ be\ positive\ definite$$
$$m(m - 2a) = 0 \to m = 2a$$

במצב זה $m > a$ ולכן ה-*trace* שלילי וקיימים ע״ע שליליים, כלומר המטריצה לא יכולה להיות מוגדרת אי שלילית.
עבור כל $m > 0$.

בנוסף, הראנו בסעיף הקודם כי ההסיאן מוגדר אי שלילית לכל x בתחום, ולכן, כל נקודה סטציונרית היא נקודת מינימום גלובלי של הפונקציה. לכן, קיים פתרון מיטבי.

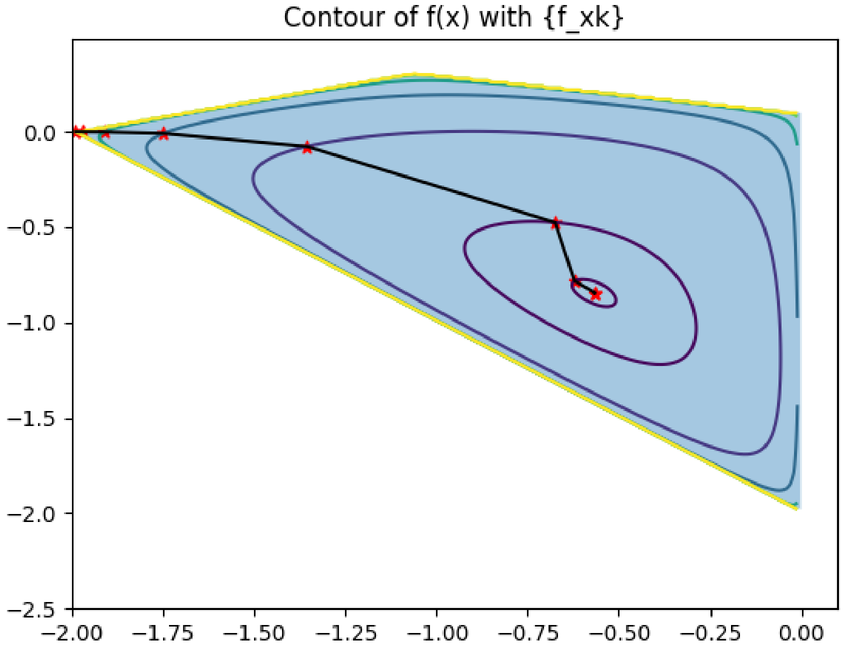
ד. *analytic_center*:

```
def armijo_newton(f, df, ddf, alpha, beta, s, xk, eps):
    xs, fs = [], []
    while np.linalg.norm(df(xk)) > eps:
        f_xk, df_xk = f(xk), df(xk)
        xs.append(xk.T)
        fs.append(f_xk)
        dk = np.linalg.solve(ddf(xk), -df_xk)
        tk = s
        while f(xk + tk * dk) >= f_xk + alpha * tk * df_xk.T.dot(dk):
            tk = beta * tk
        xk = xk + tk * dk
    if len(xs) == 0:
        xs, fs = [xk.T], [f(xk)]
    return xs, fs


def analytic_center(A, b, x0):
    if not all([(A[i].T.dot(x0) <= b[i]) for i in range(len(b))]):
        raise Exception('x0 is not in interior(P)')
    xs, fs = armijo_newton(f(A, b), df(A, b), ddf(A, b),
                           alpha=1 / 4,
                           beta=1 / 2,
                           s=2,
                           xk=x0,
                           eps=1 / np.power(10, 6))
    return np.array(xs), np.array(fs)
```

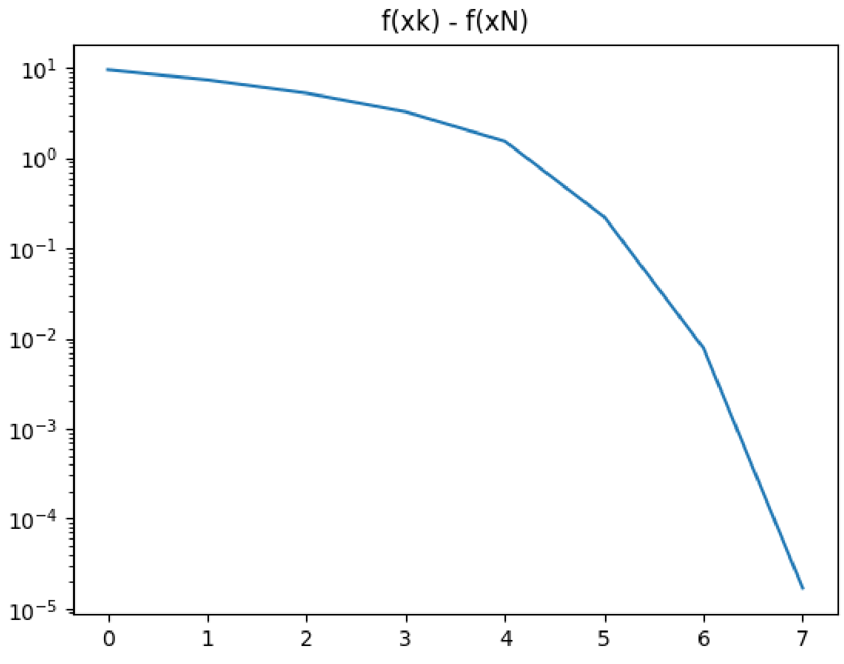
ה. *test_analytic_center*:

```
def test_analytic_center():
    A = np.array([[2, 10],
                  [1, 0],
                  [-1, 3],
                  [-1, -1]])
    b = np.array([1, 0, 2, 2])
    x0 = np.array([-1.99, 0])
    xs, fs = analytic_center(A, b, x0)

    X, Y = np.meshgrid(np.arange(-2, 0, 0.01), np.arange(-2.5, 0.5, 0.01))
    Z = X.copy()

    f_xk = f(A, b)
    for i in range(len(X)):
        for j in range(len(X[0])):
            Z[i, j] = f_xk((X[i, j], Y[i, j]))

    Z[(np.isnan(Z))] = 10 ** 10
    fig, ax = plt.subplots(1, 1)
    ax.contour(X, Y, Z, levels=fs[::-1], extend='both')
    patches = []
    polygon = Polygon([[-2, 0],
                       [-1.0604, 0.3072],
                       [-0.001, 0.1],
                       [-0.001, -1.98]], True)

    patches.append(polygon)

    p = PatchCollection(patches, alpha=0.4)
    ax.add_collection(p)
    ax.scatter(xs[:, 0], xs[:, 1], marker='*', color='red')
    ax.plot(xs[:, 0], xs[:, 1], color='black')
    plt.title("Contour of f(x) with {f_xk}")
    plt.show()

    plt.figure()
    plt.title("f(xk) - f(xN)")
    plt.semilogy(np.arange(len(fs) - 1), np.array(fs[0:len(fs) - 1]) - fs[-1], label="f(xk) -
f(xN)")
    plt.show()
```

תרשימים:



Contour of f(x) with {f_xk}
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−2.00
−1.75
−1.50
−1.25
−1.00
−0.75
−0.50
−0.25
0.00



f(xk) - f(xN)
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
0
1
2
3
4
5
6
7

שאלה 2 :

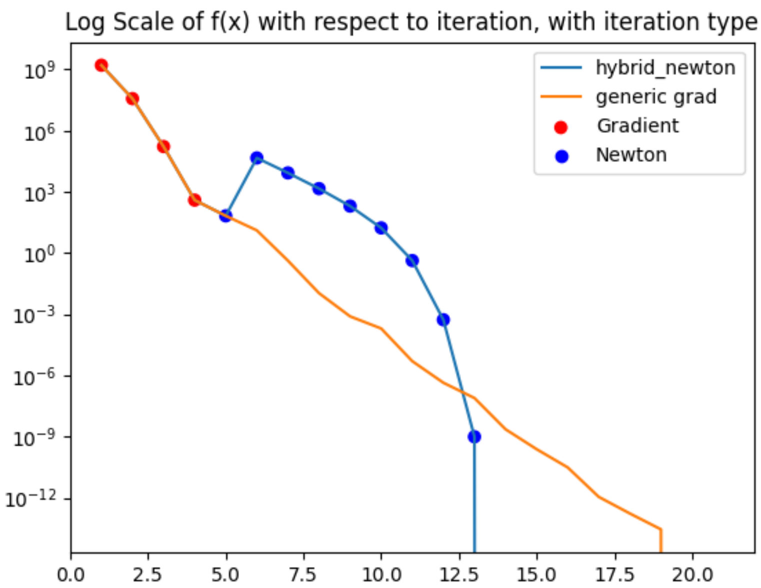
א. *hybrid_newton* :

```
def hybrid_newton(f, gf, hf, lsearch, xk, eps):
    f_xk, df_xk, hf_xk = f(xk), gf(xk), hf(xk)
    fs, gs, ts, newton = [f_xk], [np.linalg.norm(df_xk)], [0], []
    while np.linalg.norm(gf(xk)) > eps:
        start_time = time.time()
        try:
            np.linalg.cholesky(hf(xk))
            dk = np.linalg.solve(hf_xk, -df_xk)
            tk = lsearch(xk, df_xk, [dk, 'newton'])
            newton.append(1)
        except LinAlgError:
            dk = -df_xk
            tk = lsearch(xk, df_xk, [dk, 'grad'])
            newton.append(0)
        xk = xk + tk * dk
        f_xk, df_xk, hf_xk = f(xk), gf(xk), hf(xk)
        fs.append(f_xk)
        gs.append(np.linalg.norm(df_xk))
        ts.append(time.time() - start_time + ts[-1])
    newton.append(1)
    return xk, fs, gs, ts[1:], newton
```

ב. *hybrid_back* :

```
def hybrid_back(f, alpha, beta, s):
    def lsearch(xk, gk, direction):
        if direction[1] == 'newton':
            return 1
        else:
            dk = direction[0]
            tk = s
            while f(xk + tk * dk) >= f(xk) + alpha * tk * gk.T.dot(dk):
                tk *= beta
            return tk

    return lsearch
```

ג. תרשים :



האלגוריתם ההיברידי אינו אלגוריתם ירידה, כפי שניתן לראות באיטרציה החמישית עקב שיטת ניוטון ערך הפונקציה עולה.
ניתן להוסיף בדיקה שרק במידה ובאמצעות שיטת ניוטון נקבל ערך פונקציה נמוך יותר נבחר בה, אחרת נשתמש בשיטת הגרדיאנט שאנו יודעים כי היא מונוטונית יורדת.

שאלה 3 :

נתונה בעיית פרמה וובר

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \equiv \sum_{i=1}^{m} \omega_i \|\mathbf{x}-\mathbf{a}_i\|,$$

נגדיר :

$$h(x,y) = \sum_{i=1}^{m} \frac{\omega_i(\|x-a_i\|^2)}{\|y-a_i\|}$$

$$L(x) = \sum_{i=1}^{m} \frac{\omega_i}{\|x-a_i\|}$$

א. נשים לב כי עבור $h(x,y) \rightarrow x$ מתקיים כי y קבוע.

$$h(x,y) = \sum_{i=1}^{m} \frac{\omega_i(\|x-a_i\|^2)}{\|y-a_i\|} = h(x,y) = \sum_{i=1}^{m} \frac{\omega_i(x^T I x - 2x^T a_i + a_i^T a_i)}{\|y-a_i\|} =$$

$$x^T I x \sum_{(i=1)}^{m} \frac{\omega_i}{\|y-a_i\|} + \sum_{i=1}^{m} \frac{-2x^T a_i}{\|y-a_i\|} + \sum_{i=1}^{m} \frac{a_i^T a_i}{\|y-a_i\|} = x^T L(y) I x + \sum_{i=1}^{m} \frac{-2x^T a_i}{\|y-a_i\|} + \sum_{i=1}^{m} \frac{a_i^T a_i}{\|y-a_i\|}$$

ניתן לראות כי זו פונקציה ריבועית עם מטריצה מובילה $L(y)I_n$.

$$h(x,y) = h(y,y) + \frac{\partial h}{\partial x}(y,y)(x-y) + (x-y)^T \nabla^2 h(y,y)(x-y)$$

הראנו כי

$$\nabla^2 h(x,y) = L(y) I_n = L(y)$$

ולכן נקבל כי :

$$h(x,y) = h(y,y) + \frac{\partial h}{\partial x}(y,y)(x-y) + L(y)\|x-y\|^2$$

ב. ניתן לראות כי

$$h(y,y) = \sum_{i=1}^{m} \frac{\omega_i(\|y-a_i\|^2)}{\|y-a_i\|} = \sum_{i=1}^{m} \omega_i(\|y-a_i\|)$$

$$\nabla f(y) = \sum_{i=1}^{m} \omega_i \frac{y-a_i}{\|y-a_i\|}$$

$$\frac{\partial h}{\partial x}(x,y) = \sum_{i=1}^{m} 2\omega_i \cdot \frac{x-a_i}{\|y-a_i\|} \rightarrow \frac{\partial h}{\partial x}(y,y) = \sum_{i=1}^{m} 2\omega_i \cdot \frac{y-a_i}{\|y-a_i\|} = 2\nabla f(y)$$

לכן :

$$2f(x) - f(y) = 2\sum_{i=1}^{m} \omega_i(\|x-a_i\|) - \sum_{i=1}^{m} \omega_i(\|y-a_i\|) = \sum_{i=1}^{m} \omega_i(2\|x-a_i\| - \|y-a_i\|)$$

צריך להוכיח כי

$$\sum_{i=1}^{m}\frac{\omega_i(\|x-a_i\|^2)}{\|y-a_i\|}\geq\sum_{i=1}^{m}\omega_i(2\|x-a_i\|-\|y-a_i\|)$$

נראה זאת על ידי כך שנוכיח כי האיבר הכללי של הסכום השמאלי גדול או שווה לאיבר הכללי של הסכום הימני, כלומר :

$$\frac{\omega_i(\|x-a_i\|^2)}{\|y-a_i\|}\geq\omega_i(2\|x-a_i\|-\|y-a_i\|)\rightarrow\|x-a_i\|^2\geq 2\|x-a_i\|\|y-a_i\|-\|y-a_i\|^2\rightarrow$$

$$(\|x-a_i\|-\|y-a_i\|)^2\geq 0$$

אי השוויון האחרון נכון תמיד ולכן

$$h(x,y)\geq 2f(x)-f(y)$$

כעת, ניתן להסיק כי :

$$f(x)\leq\frac{h(x,y)+f(y)}{2}=\frac{f(y)+2\nabla f(y)(x-y)+L(y)\|x-y\|^2+f(y)}{2}$$
$$=f(y)+\nabla f(y)(x-y)+\frac{L(y)\|x-y\|^2}{2}$$

שאלה 4 :

עבור m פונקציות $f_i:\mathbb{R}^n\rightarrow\mathbb{R}$ גזירות ברציפות ו- m קבועים $c_i\in\mathbb{R}$.

נתונה הבעיה הבאה :

$$\min_{x\in\mathbb{R}^n}\left\{g(x)\equiv\sum_{i=1}^{m}(f_i(x)-c_i)^2\right\}$$

נגדיר $F:\mathbb{R}^n\rightarrow\mathbb{R}^m$ כך :

$$F(x)=\begin{pmatrix}(f_1(x)-c_1)\\ \vdots\\ (f_m(x)-c_m)\end{pmatrix}$$

נסמן את מטריצת היעקוביאן של F ב- $x\in\mathbb{R}^n$ $J(x)$.

א.
נשים לב כי

$$g(x)=\|F(x)\|^2\rightarrow\nabla g(x)=2\nabla F(x)^TF(x)\cdot=2J(x)^TF(x)$$

מכיוון שמטריצת הנגזרות החלקיות של F מוגדרת להיות מטריצת יעקובי.

ב. ניתן לראות כי עבור שיטת גאוס ניוטון מתקיים :

$$x^{k+1}=\min_{x\in\mathbb{R}^n}(\|F(x^k)+J(x^k)^T(x-x^k)\|^2)=\min_{x\in\mathbb{R}^n}(\|J(x^k)^Tx+F(x^k)-J(x^k)^Tx^k\|^2)$$

נגדיר :

$$b^k=-(F(x^k)-J(x^k)^Tx^k)$$

ונקבל :

$$\min_{x\in\mathbb{R}^n}(\|J(x^k)^Tx-b^k\|^2)$$

ג. נניח כי $J(x^k)$ מדרגת עמודות מלאה.
על ידי מזעור הפונקציה נקבל כי :

$$x^{k+1}=(J(x^k)J(x^k)^T)^{-1}J(x^k)b^k$$

שיטת הגראדינט המדורגת נראית כך :

$$x^{k+1}=x^k-t^kD^k\nabla g(x^k)=x^k-2t^k\cdot D^k\cdot J(x)^TF(x)$$

$$(J(x^k)J(x^k)^T)^{-1}J(x^k)b^k = (J(x^k)J(x^k)^T)^{-1}J(x^k)\left(J(x^k)^T x^k - F(x^k)\right)$$
$$= (J(x^k)J(x^k)^T)^{-1}\left(J(x^k)J(x^k)^T x^k - J(x^k)F(x^k)\right) =$$
$$x^k - (J(x^k)J(x^k)^T)^{-1} \cdot J(x^k)F(x^k)$$

נגדיר :

$$t^k = \frac{1}{2}, D^k = (J(x^k)J(x^k)^T)^{-1}$$

ונקבל את שיטת הגרדיאנט.

נשים לב כי נתון ש- $J(x^k)$ מדרגת עמודות מלאה, לכן, אין לה ע״ע אפס.
לאור זאת

$$J(x^k)J(x^k)^T \succ 0 \rightarrow (J(x^k)J(x^k)^T)^{-1} \succ 0\ \forall k$$

מש״ל.